

André Ricardo de Azevedo

A Função do Resumo na Comunicação Científica:  
Uma Comparação entre Resumos Tradicionais e  
Estruturados

Belo Horizonte

Escola de Ciência da Informação

2011

André Ricardo de Azevedo

# A Função do Resumo na Comunicação Científica: Uma Comparação entre Resumos Tradicionais e Estruturados

Dissertação apresentada ao Programa de Pós-Graduação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais como requisito parcial à obtenção de título de Mestre em Ciência da Informação.

Área de Concentração: Produção, Organização e Utilização da Informação.

Linha de Pesquisa: Organização e Uso da Informação.

Orientadora: Marlene Oliveira.

Belo Horizonte

2011

Azevedo, André Ricardo de.

A994f A Função do resumo na comunicação científica: uma comparação entre resumos tradicionais e estruturados [manuscrito] / André Ricardo de Azevedo. – 2011.

181 f. : il., enc.

Orientadora: Marlene Oliveira.

Inclui apêndice.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Bibliografia: f. 104-116.

1. Ciência da informação. – Teses. 2. Resumos – Redação. – Teses. 3. Comunicação na ciência – Indexação e resumos. – Teses. 4. Redação acadêmica. – Teses. I. Título. II. Oliveira, Marlene. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 001.814



UFMG

Universidade Federal de Minas Gerais  
Escola de Ciência da Informação  
Programa de Pós-Graduação em Ciência da Informação

FOLHA DE APROVAÇÃO

"A FUNÇÃO DO RESUMO NA COMUNICAÇÃO CIENTÍFICA: UMA COMPARAÇÃO ENTRE RESUMOS TRADICIONAIS E ESTRUTURADOS"

André Ricardo de Azevedo

Dissertação submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como parte dos requisitos à obtenção do título de "**Mestre em Ciência da Informação**", linha de pesquisa "**Organização e Uso da Informação - OUI**".

Dissertação aprovada em: 18 de novembro de 2011.

Por:


  
\_\_\_\_\_  
Profa. Dra. Marlene Oliveira Teixeira de Melo - ECI/UFMG (Orientadora)

  
\_\_\_\_\_  
Profa. Dra. Anneliese Maria Bento Gama de Carvalho - FALE/UFMG

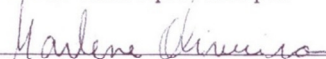
  
\_\_\_\_\_  
Profa. Dra. Adriana Bogliolo Sirihal Duarte - ECI/UFMG

  
\_\_\_\_\_  
Profa. Dra. Terezinha de Fátima Carvalho de Souza - ECI/UFMG

Aprovada pelo Colegiado do PPGCI

  
\_\_\_\_\_  
Profa. Renata Maria Abrantes Baracho Porto  
Sub-Coordenadora

Versão final Aprovada por

  
\_\_\_\_\_  
Profa. Marlene Oliveira Teixeira de Melo  
Orientadora



UFMG

**Universidade Federal de Minas Gerais**  
**Escola de Ciência da Informação**  
**Programa de Pós-Graduação em Ciência da Informação**

ATA DA DEFESA DE DISSERTAÇÃO DE **ANDRÉ RICARDO DE AZEVEDO**, matrícula:  
2009654620

Às 14:00 horas do dia 18 de novembro de 2011, reuniu-se na Escola de Ciência da Informação da UFMG a Comissão Examinadora, aprovada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação em 17/10/2011, para julgar, em exame final, o trabalho intitulado **A função do resumo na comunicação científica: uma comparação entre resumos tradicionais e estruturados**, requisito final para obtenção do Grau de MESTRE em CIÊNCIA DA INFORMAÇÃO, área de concentração: Produção, Organização e Utilização da Informação, Linha de Pesquisa: Organização e Uso da Informação - OUI. Abrindo a sessão, a Presidente da Comissão, Profa. Dra. Marlene Oliveira Teixeira de Melo, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

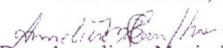
Profa. Dra. Marlene Oliveira Teixeira de Melo	APROVADO
Profa. Dra. Anneliese Maria Bento Gama de Carvalho	APROVADO
Profa. Dra. Adriana Bogliolo Sirihal Duarte	APROVADO
Profa. Dra. Terezinha de Fátima Carvalho de Souza	APROVADO

Pelas indicações, o candidato foi considerado APROVADO.

O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 18 de novembro de 2011.

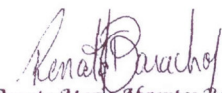
  
Profa. Dra. Marlene Oliveira Teixeira de Melo  
Orientadora - ECI/UFMG

  
Profa. Dra. Anneliese Maria Bento Gama de Carvalho  
FALE/UFMG

  
Profa. Dra. Adriana Bogliolo Sirihal Duarte  
ECI/UFMG

  
Profa. Dra. Terezinha de Fátima Carvalho de Souza  
ECI/UFMG

Obs: Este documento não terá validade sem a assinatura e carimbo da Coordenadora.

  
Profa. Renata Maria Abrantes Baracho Porto  
Sub-coordenadora do Programa de  
Pós-Graduação em Ciência da  
Informação - ECI / UFMG

## Agradecimentos

Agradeço a meus professores e colegas da graduação e da pós-graduação cujo contato e associação animaram substantivamente minha vida acadêmica e profissional e ampliaram meu olhar humano e técnico sobre essas esferas de atividade.

Agradeço à professora Marlene pela orientação pragmática, escorreita e por conversas inspiradoras.

Agradeço a meus amigos pela amizade de sempre que em nada contribuiu academicamente para a realização deste trabalho, mas cooperou com ele de um modo bizantino.

Agradeço a minha família por tornar valiosos (e tão divertidos) aspectos da minha vida distintos do meu trabalho e da minha paixão. Sou grato, em particular, a minha mãe e a meus irmãos por constituírem-se em minha noção de bem-estar e felicidade. Sou grato a meu pai por recomeçar com fé, humildade e amor.

Por Fernanda também sou muito grato.

## Resumo tradicional

Procurou-se comparar resumos de artigos científicos do Journal of Documentation (JD) e do Journal of the American Society for Information Science and Technology (JASIST), avaliando o desempenho desses resumos no que diz respeito à representação de conteúdo relevante dos textos originais que condensam. O JASIST publica junto aos artigos científicos resumos informativos/indicativos. O JD publica resumos estruturados junto a seus artigos.

Argumenta-se que a sociedade da informação é a sociedade do conhecimento teórico codificado e que a comunicação científica é um dos principais processos sobre os quais essa sociedade se assenta. Defende-se que a recuperação da informação como instrumento excepcional dentro do processo de comunicação científica vai avançar na medida em que os produtos da atividade documentária avançarem conceitualmente. Em particular, a normalização da atividade de redação de resumos precisa evoluir quanto às instruções e técnicas para seleção de conteúdos (e não só para sua apresentação).

São comparados 42 resumos informativos/indicativos e 42 resumos estruturados. Primeiramente, comparou-se o resumo com o artigo científico que condensa. Para isso, foram usados dois conceitos: superestrutura textual (conceito tornado operacional por proposta aplicada nesta pesquisa) e estrutura temática (conceito analisado subjetivamente). A superestrutura textual é reconhecida por uma 'sequência de base' como por exemplo: problema, hipótese, metodologia, resultados, conclusão. A estrutura temática é que faz reconhecer o assunto de um texto e estabelece a 'similaridade textual' – no nível da mensagem. A superestrutura textual divide o texto em partes articuladas e estabelece a 'contiguidade textual' – no nível da estrutura, que evidencia as relações estabelecidas.

A análise, com poder estatístico de aproximadamente 0.79 (usando um teste qui-quadrado com  $GL=1$ ), encontrou que 3/42 ou 7,14% dos resumos informativos proporcionam uma representação satisfatória em relação ao artigo científico original, e, 14/42 ou 33,3% dos resumos estruturados proporcionam uma representação satisfatória em relação ao artigo científico original. O uso da contiguidade textual para analisar o texto original e o do resumo se mostrou prático, e, um resumo estruturalmente semelhante ao seu original mostrou proporcionar uma representação 'interpretativa' eficaz.

## Resumo estruturado

### **Propósito**

Procurou-se comparar resumos de artigos científicos do Journal of Documentation (JD) e do Journal of the American Society for Information Science and Technology (JASIST), avaliando o desempenho desses resumos no que diz respeito à representação de conteúdo relevante dos textos originais que condensam. O JASIST publica junto aos artigos científicos resumos informativos/indicativos. O JD publica resumos estruturados junto a seus artigos.

### **Abordagem/Metodologia**

Argumenta-se que a sociedade da informação é a sociedade do conhecimento teórico codificado e que a comunicação científica é um dos principais processos sobre os quais essa sociedade se assenta. Defende-se que a recuperação da informação como instrumento excepcional dentro do processo de comunicação científica vai avançar na medida em que os produtos da atividade documentária avançarem conceitualmente. Em particular, a normalização da atividade de redação de resumos precisa evoluir quanto às instruções e técnicas para seleção de conteúdos (e não só para sua apresentação).

São comparados 42 resumos informativos/indicativos e 42 resumos estruturados. Primeiramente, comparou-se o resumo com o artigo científico que condensa. Para isso, foram usados dois conceitos: superestrutura textual (conceito tornado operacional por proposta aplicada nesta pesquisa) e estrutura temática (conceito analisado subjetivamente). A superestrutura textual é reconhecida por uma 'sequência de base' como por exemplo: problema, hipótese, metodologia, resultados, conclusão. A estrutura temática é que faz reconhecer o assunto de um texto e estabelece a 'similaridade textual' – no nível da mensagem. A superestrutura textual divide o texto em partes articuladas e estabelece a 'contiguidade textual' – no nível da estrutura, que evidencia as relações estabelecidas.

### **Resultados**

A análise, com poder estatístico de aproximadamente 0.79 (usando um teste qui-quadrado com  $GL=1$ ), encontrou que 3/42 ou 7,14% dos resumos informativos proporcionam uma representação satisfatória em relação ao artigo científico original, e, 14/42 ou 33,3% dos resumos estruturados proporcionam uma representação satisfatória em relação ao artigo científico original. O uso da contiguidade textual para analisar o texto original e o do resumo se mostrou prático, e, um resumo estruturalmente semelhante ao seu original mostrou proporcionar uma representação 'interpretativa' eficaz.

### **Originalidade e Valor**

Contribui para o diminuto debate nacional sobre a questão do resumo em comunicação científica. Propõe um avanço no processo de redação de resumos dentro desse contexto, usando uma proposta cujo valor vem sendo comprovado fora do Brasil. Problematiza o uso de resumos indicativos no processo de recuperação da informação sugerindo o seu abandono. Apresenta o tratamento da informação como uma atividade que possui um caráter eminentemente epistemológico e que deve avançar como uma atividade desse tipo.

## Abstract

We sought to compare abstracts of scientific articles in the Journal of Documentation (JD) and the Journal of the American Society for Information Science and Technology (JASIST), evaluating the performance of these abstracts with regard to the representation of relevant content from the original texts. The JASIST publishes along with its scientific papers traditional abstracts. The JD publishes structured abstracts.

It is argued that the information society is the society of coded and theoretical knowledge and that scientific communication is one of the main processes on which this society rests. It is also argued that the information retrieval, as an exceptional instrument in the process of scientific communication, will advance to the extent that abstracting and indexing advance conceptually. In particular, the normalization of abstracting must evolve in regard to instructions and techniques for content selection (and not only for its presentation).

We compared 42 traditional abstracts and 42 structured abstracts. First, the abstract was compared to the scientific article that it condenses. It was used two concepts: textual superstructure (a concept made operational by a proposal applied in this study) and thematic structure (concept analyzed subjectively). The textual superstructure is recognized by a 'base sequence' such as: problem, hypothesis, methodology, results, conclusion. The thematic structure is the "subject" of a text. This structure establishes the 'textual similarity' - at the message level. The textual superstructure divides the text into articulated parts and establishes the 'textual contiguity' - at the structure level (which shows the relationships established).

The analysis with statistical power of approximately 0.79 (using a chi-square with DF=1) found that 3/42 or 7.14% of the traditional abstracts provide a satisfactory representation in relation to the original scientific article, and, 14/42 or 33.3% of the structured abstracts provide a satisfactory representation in relation to the original scientific article. The use of textual contiguity to analyze the original text of the article and its abstract was practical. An abstract structurally similar to its original provides an effective 'interpretive' representation.

## Lista de Abreviaturas

ABNT	Associação Brasileira de Normas Técnicas
ASIST	American Society for Information Science and Technology
CAPES	Coordenação para Aperfeiçoamento de Pessoal de Ensino Superior
ISO	International Organization for Standardization
JASIST	Journal of the American Society for Information Science and Technology
JD	Journal of Documentation
NBR	Norma Brasileira

## Sumário

	Introdução .....	1
1	Problema de pesquisa .....	4
1.1	Justificativa .....	7
1.2	Objetivo Geral .....	10
1.3	Objetivos Específicos .....	10
2	Referencial teórico .....	11
2.1	A sociedade da informação .....	11
2.2	A comunicação científica .....	18
2.3	A recuperação da informação científica .....	32
2.4	O resumo na comunicação científica .....	41
2.5	A elaboração de resumos .....	56
3	Metodologia .....	71
3.1	Corpus da pesquisa .....	75
3.2	Procedimentos de coleta dos dados .....	76
3.3	Procedimentos de análise dos dados .....	77
4	Resultados .....	79
5	Considerações finais .....	94
6	Sugestão de Pesquisas .....	102
	Referências bibliográficas .....	104
	Apêndice A Consulta a pesquisadores em ciência da informação e engenharia	117
	Apêndice B Corpus analisado nesta pesquisa.....	129

# Introdução

Os conhecimentos que uma sociedade encerra a orientam em suas escolhas e na determinação do seu modo de vida. Os conhecimentos compõem nosso sistema explicativo da realidade e instruem nossa ação no mundo, fundamentalmente, por meio de crenças e pressupostos.

A atividade de conhecer pode ser dirigida por diferentes referenciais epistemológicos: a intuição, a tradição, a autoridade, o saber racional (LAVILLE e DIONNE, 1999). Na maioria das sociedades contemporâneas o saber racional – nas feições da ciência moderna – adquiriu, em relação aos referenciais concorrentes, um status de saber privilegiado, principalmente, devido a sua efetividade em modelar diversos aspectos da vida social.

A velocidade do debate e do desenvolvimento científico, entretanto, vem provando ser um desafio para pesquisadores e profissionais enquanto agentes participativos em suas respectivas áreas de atuação e em sua capacidade de influenciarem de maneira informada o processo de construção social. O volume de informações produzido globalmente, mesmo em uma área especializada do conhecimento, é prodigioso e exige do sujeito que tenciona conhecer uma estratégia sofisticada para o uso eficiente de sua capacidade intelectual.

O estudo da comunicação científica, nesse sentido, oferece princípios e conceitos importantes para a atuação do profissional nesse contexto dinâmico. Além disso, o estudo da comunicação científica é valioso para o próprio entendimento do que é a ciência. Conforme Alvarenga (2000, p. 123) “... resultados de pesquisas, que têm por objeto a comunicação científica, podem ser vistos como subsídios para os estudos epistemológicos e arqueológicos no conhecimento de campos disciplinares constantes do sistema dos saberes.”

Especificamente, o estudo dos resumos e do processo de condensação da literatura científica permite uma melhor compreensão de aspectos essenciais desse processo que põe em contato toda a comunidade científica – aspectos ligados à cognição e ao próprio processo comunicativo. Resumos de artigos científicos, também, são instrumentos estratégicos para o gerenciamento automático e não-automático, rápido e acurado de grandes volumes de informação e permitem o estabelecimento de associação e relações entre diferentes conteúdos visando à geração de novos conhecimentos. Presentemente, no entanto, o resumo vem apresentando uma qualidade contingente enquanto item documentário.

Os resumos de uso mais difundido no meio técnico-científico são do tipo informativo e indicativo. No entanto, outro tipo de resumo – o resumo estruturado – foi proposto no final da década de 1980 e, desde então, tem se mostrado um instrumento informativo valioso dentro do contexto atual.

A presente pesquisa se propõe a examinar mais de perto o uso desses tipos de resumo na comunicação científica.

No capítulo 1 introduz-se o problema de pesquisa. É discutida sua validade e são traçados objetivos para se verificar em que medida a solução proposta no final deste texto e o problema formulado se ajustam.

No capítulo 2 apresenta-se o referencial teórico.

Na primeira parte do referencial teórico, é feita uma discussão sobre a sociedade da informação. O intuito dessa discussão é situar ciência e comunicação científica em um contexto social onde não é excessivamente valorizada a influência de fatores como o tecnológico ou o econômico, por exemplo. Esse tipo de sobrevalorização acontece nas análises propostas por vários teóricos da sociedade da informação. Na discussão aqui feita, defende-se que a sociedade da informação é fortemente marcada pelo *conhecimento teórico*, e esse fator é determinante em sua identidade.

Na segunda parte, discute-se aspectos ligados à comunicação científica – processo que sustenta e estabiliza a chamada sociedade da informação (e do conhecimento). É feita uma apresentação dos aspectos social e técnico do periódico científico e ressalta-se a

importância da codificação do conhecimento por meio desse instrumento. Analisam-se as características dos canais formais e informais da comunicação científica, e o papel que a avaliação pelos pares (peer review) desempenha nessa caracterização. Discute-se ainda a comunicação científica a partir do ciclo informacional e a questão de como o resumo aumenta a acessibilidade de um documento.

Na terceira parte, são examinados aspectos da recuperação da informação científica e são definidos conceitos ligados à recuperação e ao tratamento da informação. É problematizado o modelo teórico que orienta a recuperação da informação e é defendida a ideia de que o tratamento da informação é uma atividade de caráter epistemológico.

Na quarta parte, debate-se o resumo como item documentário no processo de comunicação científica. Apresenta-se seu histórico e sua tipologia. São discutidas as normas ISO 214 e NBR 6028. Examina-se o resumo informativo em comparação com o estruturado e é feita uma revisão da literatura sobre resumos.

Na quinta parte, aborda-se a elaboração de resumos. É apresentada a proposta de Kobashi (1997) sobre a superestrutura textual e estrutura temática de certos tipos de texto. Essa superestrutura vai repercutir na elaboração de resumos e fundamenta a análise dos resultados desta pesquisa. Nessa parte, é problematizado o uso de resumos indicativos na comunicação científica. Também, são apresentadas algumas tentativas de sistematização da atividade de elaboração de resumos no nível do conteúdo informativo. É discutida ainda a textualidade do resumo e aspectos da interdisciplinaridade que implica.

No capítulo 3 é explicada a metodologia usada, a definição do corpus da pesquisa, e os procedimentos de coleta e análise dos dados. No capítulo 4 são apresentados e discutidos os resultados. No capítulo 5 são feitas algumas considerações como epílogo. No último capítulo são feitas algumas sugestões para futuras pesquisas. No Apêndice A é relatada uma consulta feita a pesquisadores em ciência da informação e engenharia, e no Apêndice B são listados os resumos analisados aqui.

# 1. Problema de pesquisa

Define-se resumo informativo como o resumo que *informa* o conteúdo de um documento, repetindo a parte predicativa desse conteúdo. Esse tipo de resumo apresenta efetivamente dados quantitativos e qualitativos presentes no texto original. O resumo informativo comunica ideias principais, resultados e conclusões de forma a poupar a consulta ao texto integral.

Um exemplo de resumo informativo poderia ser:

SAVOLAINEN, Reijo. Critérios de preferência de fonte no contexto de projetos do cotidiano: julgamentos de relevância feitos por indivíduos interessados na compra de imóveis. **Journal of Documentation**, v. 66, n. 1, p. 70-92, 2010.

O propósito deste artigo é explorar como critérios de preferência de fonte são definidos no contexto de projetos do cotidiano que requerem busca de informação para problemas específicos. Mais especificamente, descobrir como aqueles que buscam por informação explicam seus critérios de preferência, caracterizando as forças e fraquezas percebidas nas diferentes fontes. A abordagem toma a forma de uma análise qualitativa de conteúdo de dados empíricos reunidos por entrevistas semi-estruturadas com 16 interessados na compra de imóveis em 2008. Os critérios de preferência de fonte foram levantados por meio da construção do horizonte das fontes de informação. Fontes envolvendo rede de contatos foram mais favorecidas, seguidas por mídia impressa, fontes humanas e fontes organizacionais. O conteúdo da informação foi o critério primário de preferência de fonte. Disponibilidade da informação foi um critério bastante importante, enquanto que características do usuário, usabilidade da informação e fatores situacionais foram bastante marginais. Na definição de critérios de preferência, mais ênfase foi colocada nas forças do que nas fraquezas percebidas nas fontes. Qualidades positivas como “provê informação atualizada” foram referidas, particularmente enquanto se julgava a relevância das fontes que envolviam rede de contatos. Qualidades negativas como “informação desatualizada” foram primeiramente associadas com mídia impressa e fontes organizacionais. O estudo é exploratório, baseando-se numa amostra relativamente pequena recrutada através de um serviço disponível na web. Dessa forma, os resultados não podem ser generalizados a potenciais interessados na compra de imóveis. Interessados na compra de imóveis tendem a favorecer fontes e serviços disponíveis na web. Esses serviços deveriam prover os clientes com informação detalhada sobre o imóvel, incluindo fotos. O artigo especifica o modelo de julgamento de relevância usado pelo usuário no contexto da busca por informação no dia-a-dia.

O Resumo Indicativo *indica* que informação significativa pode ser encontrada no documento original sem dar conta de que informação é essa. Ele trata das características temáticas mais gerais do documento sem comunicar, por exemplo, os resultados ou a metodologia relatados no texto original. A proposta desse formato de resumo é a de recomendar a leitura do original e, de forma nenhuma, lhe poupar a consulta.

Um exemplo de resumo indicativo poderia ser:

SAVOLAINEN, Reijo. Critérios de preferência de fonte no contexto de projetos do cotidiano: julgamentos de relevância feitos por indivíduos interessados na compra de imóveis. **Journal of Documentation**, v. 66, n. 1, p. 70-92, 2010.

Trata de critérios de preferência de fonte de informação definidos no contexto de projetos do cotidiano que requerem busca de informação para problemas específicos. Critérios de preferência de fonte foram levantados por meio da construção do horizonte das fontes de informação. Na definição de critérios de preferência, mais ênfase foi colocada nas forças do que nas fraquezas percebidas nas fontes. Interessados na compra de imóveis tendem a favorecer fontes e serviços disponíveis na web. O estudo é exploratório e os resultados obtidos a partir de um corpus pequeno e não-aleatório não devem ser generalizados. O artigo especifica o modelo de julgamento de relevância usado pelo usuário no contexto da busca por informação no dia-a-dia.

O Resumo Estruturado é um resumo informativo que estrutura seu conteúdo em parágrafos sob cabeçalhos como: Antecedentes, Objetivo, Método, Resultados, Conclusão.

Um exemplo de resumo estruturado poderia ser:

SAVOLAINEN, Reijo. Critérios de preferência de fonte no contexto de projetos do cotidiano: julgamentos de relevância feitos por indivíduos interessados na compra de imóveis. *Journal of Documentation*, v. 66, n. 1, p. 70-92, 2010.

**Propósito** – O propósito deste artigo é explorar como critérios de preferência de fonte são definidos no contexto de projetos do cotidiano que requerem busca de informação para problemas específicos. Mais especificamente, descobrir como aqueles que buscam por informação explicam seus critérios de preferência, caracterizando as forças e fraquezas percebidas nas diferentes fontes.

**Metodologia/abordagem** – A abordagem toma a forma de uma análise qualitativa de conteúdo de dados empíricos reunidos por entrevistas semiestruturadas com 16 interessados na compra de imóveis em 2008. Os critérios de preferência de fonte foram levantados por meio da construção do horizonte das fontes de informação.

**Resultados** – Fontes envolvendo rede de contatos foram mais favorecidas, seguidas por mídia impressa, fontes humanas e fontes organizacionais. O conteúdo da informação foi o critério primário de preferência de fonte. Disponibilidade da informação foi um critério bastante importante, enquanto que características do usuário, usabilidade da informação e fatores situacionais foram bastante marginais nesse respeito. Na definição de critérios de preferência, mais ênfase foi colocada nas forças do que nas fraquezas percebidas nas fontes. Qualidades positivas como “provê informação atualizada” foram referidas, particularmente enquanto se julgava a relevância das fontes que envolviam rede de contatos. Qualidades negativas como “informação desatualizada” foram primeiramente associadas com mídia impressa e fontes organizacionais.

**Limitações/implicações da pesquisa** – O estudo é exploratório, baseando-se numa amostra relativamente pequena recrutada através de um serviço disponível na web. Dessa forma, os resultados não podem ser generalizados a potenciais interessados na compra de imóveis.

**Implicações práticas** – Interessados na compra de imóveis tendem a favorecer fontes e serviços disponíveis na web. Esses serviços deveriam prover os clientes com informação detalhada sobre o imóvel, incluindo fotos.

**Originalidade/valor** – O artigo especifica o modelo de julgamento de relevância usado pelo usuário no contexto da busca por informação no dia-a-dia.

**Tipo de artigo** – Pesquisa.

A pergunta que a presente pesquisa pretende responder é a seguinte:

Os resumos estruturados apresentam, em relação aos resumos informativos, um maior grau de similaridade com os textos originais a partir dos quais foram produzidos, tornando-se produtos documentários que veiculam mais informação relevante?

## 1.1 Justificativa

A necessidade por melhores métodos no processo de comunicação científica vem sendo sentida de forma cada vez mais aguda à medida que mais e mais informações, em todas as frentes de atuação da ciência, são produzidas e disseminadas. Isso é especialmente verdadeiro por tratar-se de um processo que é muito complexo e com uma alta entropia.

Donald e Ana Cleveland (2001) observam que

“A necessidade futura para a redação de resumos críticos é muito forte com a sobrecarga de informação na Internet. Há uma necessidade de métodos novos e mais sofisticados de filtragem que separe informação de qualidade de inutilidades em todos os níveis.”  
(CLEVELAND e CLEVELAND, 2001, p. 56).

Roy Pitkin, Mary Ann Branagan e Leon Burmeister, em um artigo publicado no JAMA, em 1999, apontam evidências de que a proporção de resumos que apresentam algum tipo de deficiência, em relação ao texto do qual são sucedâneos, varia entre 18% e 68% (PITKIN *et al.*, 1999). A variação observada é relativa a um total de seis periódicos examinados dentro da área médica como indica a TAB. 1 abaixo.

**TABELA 1**  
Resumos deficientes

Periódico	% Deficiente (Intervalo de Conf. 95%)	Deficiências		
		Inconsistência	Omissão	Ambos
A	18 (6-30)	2	2	4
B	43 (29-58)	12	5	2
C	30 (16-43)	6	4	3
D	45 (30-59)	9	7	4
E	32 (18-45)	6	7	1
F	68 (54-82)	15	4	11

Fonte: Adaptado de PITKIN *et al.*, 1999, p. 1111.

Essa proporção de desacertos é inquietante, especialmente em um meio caracterizado por rigor e método como o meio científico. Esses resultados caracterizam um verdadeiro distúrbio informacional no processo de comunicação científica, na verdade, eles implicam que considerável volume de informação tem o seu acesso comprometido justamente por causa de itens documentários que ‘discordam’ dos documentos originais que deveriam representar.

De forma semelhante, Hahs-Vaughn e Onwuegbuzie (2010), analisando os resumos de um periódico na área de educação, constataram que esses resumos apresentavam omissões e inconsistências conforme a tabela abaixo.

**TABELA 2**  
Frequência de omissões nos resumos, por componentes

Componente do Resumo	%
Problema/Propósito/Objetivo/Pergunta de Pesquisa/Foco do Estudo	31,4
Amostra/Tamanho da População/Características	51,4
Método (procedimento de coleta de dados, intervenção, desenho da pesquisa)	8,6
Achados	17,1
Conclusões/Implicações/Recomendações	5,7

Fonte: Adaptado de HAHS-VAUGHN e ONWUEGBUZIE, 2010, p. 57.

Os resultados da TAB. 2 foram encontrados com base na análise de 74 resumos e artigos publicados em dois anos. Foi identificado que 44,3% dos resumos foram mal-desenvolvidos, deixando de comunicar informação relevante sobre algum dos “componentes” da superestrutura textual.

Hahs-Vaughn e Onwuegbuzie (2010) argumentam que o atual formato dos resumos de artigos científicos em educação (o formato informativo/indicativo) é uma barreira para a disseminação e o uso da informação. Eles sugerem que o resumo estruturado pode ajudar para uma disseminação mais efetiva e eficiente.

Outro aspecto estudado por Hahs-Vaughn e Onwuegbuzie (2010) diz respeito ao processo de submissão de artigos para publicação. Eles verificaram que artigos submetidos para publicação, com seus respectivos resumos mal-desenvolvidos, têm quase o dobro de chance de serem rejeitados em relação aos artigos com resumos bem desenvolvidos.

A legibilidade dos resumos é ainda outro aspecto a ser considerado na discussão a respeito de aperfeiçoamento no processo de condensação informacional. Autores como G. B. Dronberger e G. T. Kowitz, e, R. King avaliaram a legibilidade de resumos em publicações e chegaram a resultados convergentes. Dronberger e Kowitz (1975) verificaram que os resumos do periódico *Research in Education* apresentavam níveis de legibilidade muito baixos. King (1976) apurou que os resumos do periódico *Child Development Abstracts* eram de legibilidade insatisfatória.

A má qualidade na composição textual em um resumo compromete sua função como item documentário. Um resumo redigido em um parágrafo único tende a suscitar uma leitura algo indigesta e é mais suscetível de terminar amontoando informações sem articulá-las bem.

A correção das falhas aparentes no processo de recuperação da informação científica depende de mais esforços de pesquisa. A pesquisa aqui proposta pretende fazer uma contribuição nesse sentido.

## 1.2 Objetivo Geral

Comparar resumos de artigos científicos de dois periódicos em ciência da informação, avaliando o desempenho dos resumos informativos em relação ao desempenho dos resumos estruturados no que diz respeito à representação de conteúdo relevante dos textos originais que condensam. Um dos periódicos usados na comparação (o JASIST) reproduz resumos do tipo informativo/indicativo junto aos artigos que publica e o outro (o JD) reproduz resumos estruturados junto aos seus artigos.

## 1.3 Objetivos Específicos

- 1) A partir da leitura de cada artigo científico do corpus da pesquisa, depreender seu conteúdo e identificar sua superestrutura textual (ex.: problema, hipótese, metodologia, resultados, conclusão).
- 2) Verificar se conteúdo relevante de cada item da superestrutura textual empregada em um determinado artigo científico foi contemplado no resumo desse artigo.
- 3) Confrontar o desempenho dos resumos informativos e o desempenho dos resumos estruturados quanto à similaridade e contiguidade textuais. A similaridade textual é verificada no nível da mensagem (conteúdo), e a contiguidade textual é verificada no nível da estrutura usada para transmitir aquela mensagem. Para os fins desta pesquisa, similaridade e contiguidade textuais determinam informação relevante a ser veiculada.

## 2. Referencial Teórico

### 2.1 A sociedade da informação

O epíteto “sociedade da informação” possui um grande valor qualificativo, mas, um reduzido efeito elucidativo. Destacados estudiosos defendem a tese da centralidade da informação nos atuais fenômenos sociais, políticos e econômicos, mas, o modo como entendem que características demarcam essa centralidade são bastante diferentes.

Em suas obras mais influentes a respeito da “era da informação”, Daniel Bell (1978) avança uma concepção de sociedade pós-industrial na qual a informação cumpre um papel especial; Manuel Castells (2007) estrutura uma visão de um sistema capitalista informacional que opera de forma generalizada na sociedade em rede; Anthony Giddens (1987, 2002) se preocupa com o que ele chama de *modernização reflexiva* – processo próprio da sociedade moderna e caracterizado por uma forte presença do conhecimento e da reflexão na vida social – e com aspectos do uso da informação para fins de controle e vigilância.

Frank Webster em sua obra “Teorias da Sociedade da Informação” (*Theories of the Information Society*) discute detalhadamente as teorias desses pensadores, dentre outros. Na análise que faz em sua obra, Webster mostra-se particularmente interessado em demonstrar as repercussões que as diferentes definições de sociedade da informação vão ter nos sistemas teóricos propostos com base nelas. Webster apresenta e discute essas repercussões tanto nas teorias daqueles pensadores que anunciam o caráter revolucionário e novo deste momento histórico – fortemente marcado pelos efeitos da geração e circulação de informações – quanto na teoria daqueles pensadores que reconhecem a “força do passado pesando nos desenvolvimentos de hoje” (WEBSTER, 2006, p. 8).

O modelo de análise adotado por Webster (2006) nos interessa aqui porque nos permite situar ciência e conhecimento em um contexto que não enfatiza aspectos tecnológicos ou econômicos ou ocupacionais da sociedade da informação. Tais aspectos são excessivamente valorizados por outros teóricos que abordam o assunto. Alguns acreditam que é o aspecto tecnológico que define a sociedade da informação, outros acreditam que é o aspecto econômico da informação que desempenha esse papel.

Webster (2006) argumenta que o fundamento que nos permite designar a sociedade atual como sociedade da informação é a predominante influência que o *conhecimento teórico* tem na organização social. Tal abordagem nos permite colocar em perspectiva o importante papel que o conhecimento e a ciência desempenham na sociedade atualmente e, na verdade, desempenham há muito tempo.

Webster (2006) divide as definições dos principais autores que trabalharam o tema da sociedade da informação de acordo com os critérios que eles elegem para a explicação do fenômeno informacional. Os critérios mais comuns utilizados por esses autores são: tecnológicos, econômicos, ocupacionais e espaciais. As definições de sociedade da informação propostas segundo esses critérios compartilham o pressuposto de que mudanças quantitativas ligadas à informação estão provocando o aparecimento de um sistema social qualitativamente novo.

Webster (2006) propõe que uma definição, se pretende designar as características mais salientes da sociedade da informação, não deve observar critérios quantitativos, mas, deve observar um critério mais distintivo: o conhecimento teórico, o qual parece ocupar uma posição preeminente na atualidade.

Segundo Webster (2006), na sociedade da informação os assuntos são organizados de tal maneira que a teoria é priorizada. Todas as inovações que beneficiam a sociedade são desenvolvidas a partir de princípios conhecidos e bem organizados. Nas aplicações tecnológicas, a teoria aparece como pré-requisito, orientando e instruindo a atividade produtiva.

A designação “conhecimento teórico” usada por Webster (2006) descreve toda atitude epistêmica que envolve os processos de abstração, generalização e codificação. Esses processos permitem a manipulação de realidades concretas, a nível intelectual; a extensão intuitiva de propriedades observadas no experimento com tais realidades a outras entidades de alguma forma semelhante a elas; e, a formalização do conhecimento assim gerado.

Um exemplo da importância do conhecimento teórico para a sociedade atual pode ser encontrado nos debates ambientais, por exemplo. Esses debates são, rotineiramente, informados por projeções teóricas: o crescimento da população, as condições da camada de ozônio, o volume de gases liberados na atmosfera, etc. e fica claro que as políticas práticas implementadas em uma determinada conjuntura só podem ser concebidas com base em algum tipo de modelo teórico.

Webster (2006) observa que tais modelos podem ser algumas vezes incipientes e grosseiros, mas:

“... eles e outros semelhantes nos ajudam a perceber que, porquanto o conhecimento teórico não tenha que ser ‘verdadeiro’ em nenhum sentido absoluto, ele efetivamente toma parte decisiva em nossas vidas. O conhecimento teórico usado é, frequentemente, impreciso, mas, isso não ataca o argumento de que ele é um pré-requisito para a ação.” (WEBSTER, 2006, p. 55).

De forma geral, as noções de conhecimento e informação se relacionam de uma forma muito próxima. Anthony C. Foskett (1998, p. 3) propõe: conhecimento é o que *eu* sei, informação é o que *nós* sabemos, isto é, conhecimento compartilhado. É até natural que a noção de informação tenha sobrepujado a de conhecimento na ‘sociedade da informação’. Os efeitos que os chamados produtos e serviços de informação – texto, gráfico, imagem e vídeo – têm sobre nossos sentidos são espetaculares. A facilidade que esses produtos e serviços proporcionam para nossos processos cognitivos é enorme, pois eles oferecem dados tratados, eliminando o esforço necessário para

extração de informação de dados brutos, a partir de observações feitas em situação real.

As operações lógicas e criativas que a atividade de conhecer envolve são realizadas a partir daquilo que chamamos *informação*, daí o apelo que a informação encerra na sociedade atual. Contudo, pode-se argumentar que a informação não existe como um ente acabado, como um ente em si no mundo real (RENDÓN ROJAS, 2005). Ela existe, efetivamente, no instante em que o ser humano produz um vínculo entre as realidades objetiva (dado) e intelectual (conhecimento). Essa operação é uma operação cognitiva e é o conhecimento que instrumentaliza a produção e o estabelecimento desse vínculo. A síntese que esse vínculo proporciona é, então, incorporada à própria substância do conhecimento, crescendo-o.

Os repositórios de conhecimento, mantidos em sociedade, nos ajudam a entender nossa situação no mundo. Em nossa história evolutiva, nós nos afastamos das determinações da natureza e da tradição; constituímos grupos e indivíduos que possuem escolha no que diz respeito ao próprio destino. Essa nova atitude desenvolvida é caracterizada por reflexividade e interrogação, por abertura a novas ideias, informações e teorias.

Uma sociedade que tem o próprio destino nas mãos, sofisticada em seu modo de vida, e exigente quanto ao seu bem-estar e conforto demanda conhecimento para explorar suas opções de ação. A geração e disponibilização de conhecimento especializado para o atendimento de demandas específicas propiciam o aparecimento de um ambiente informacional complexo e diversificado.

Giddens (2002), discutindo a influência do conhecimento na modernidade, deixa claro a complexidade do atual ambiente informacional:

“A modernidade é uma ordem pós-tradicional, mas não uma ordem em que as certezas da tradição e do hábito tenham sido substituídas pela certeza do conhecimento racional. A dúvida, característica generalizada da razão crítica moderna, permeia a vida cotidiana como a consciência filosófica, e constitui uma dimensão existencial geral do mundo social contemporâneo. A modernidade

institucionaliza o princípio da dúvida radical e insiste em que todo conhecimento tome a forma de hipótese – afirmações que bem podem ser verdadeiras, mas que por princípio estão sempre abertas à revisão e podem ter que ser, em algum momento, abandonadas. Sistemas de conhecimento acumulado – importantes influências de desençaixe [mecanismo que desloca determinadas relações sociais de seus lugares específicos, recombina-as através de grandes distâncias no tempo e no espaço] – representam múltiplas fontes de autoridade, muitas vezes contestadas internamente e divergentes em suas implicações. Nas situações a que chamo de modernidade ‘alta’ ou ‘tardia’ – nosso mundo de hoje – o eu, como os contextos institucionais mais amplos em que existe, tem que ser construído reflexivamente. Mas essa tarefa deve ser realizada em meio a uma enigmática diversidade de opções e possibilidades.” (GIDDENS, 2002, p. 10-11).

Segundo Giddens (2002), em circunstâncias de incerteza e onde múltiplas escolhas se apresentam aos indivíduos, as noções de confiança e risco têm uma aplicação particular. A modernidade, para Giddens, é uma cultura do risco. O conceito de ‘risco’ se tornou fundamental para a maneira como indivíduos de forma geral organizam a vida social. Na modernidade, o futuro é continuamente antecipado por meio da organização reflexiva dos “ambientes de conhecimento”. A modernidade procura reduzir o risco de todas as áreas da vida social.

“Na ordem pós-tradicional da modernidade, e contra o pano de fundo de novas formas de experiência mediada, a auto-identidade se torna um empreendimento reflexivamente organizado. O projeto reflexivo do eu, que consiste em manter narrativas biográficas coerentes, embora continuamente revisadas, tem lugar no contexto de múltipla escolha filtrada por sistemas abstratos. Na vida social moderna, a noção de estilo de vida assume um significado particular. Quanto mais a tradição perde seu domínio, e quanto mais a vida diária é reconstituída em termos do jogo dialético entre o local e o

global, tanto mais os indivíduos são forçados a escolher um estilo de vida a partir de uma diversidade de opções. Certamente existem também influências padronizadoras – particularmente na forma da criação da mercadoria, pois a produção e a distribuição capitalistas são componentes centrais das instituições da modernidade. No entanto, por causa da ‘abertura’ da vida social de hoje, com a pluralização dos contextos de ação e a diversidade de ‘autoridades’, a escolha de estilo de vida é cada vez mais importante na constituição da auto-identidade e da atividade diária. *O planejamento de vida reflexivamente organizado, que normalmente pressupõe a consideração de riscos filtrados pelo contato com o conhecimento especializado, torna-se uma característica central da estruturação da auto-identidade.*” (GIDDENS, 2002, p. 13, o grifo é nosso).

Autores como Roland Omnès (1996), Nico Stehr (1994), e o próprio Daniel Bell (1978) compartilham a crença de que o conhecimento teórico está difundido em todos os aspectos da vida contemporânea.

Bell (1978), por exemplo, em sua análise da sociedade pós-industrial, enfatiza que o caráter dessa nova sociedade aparece indelével na transição ocorrida do processo manufatureiro para a prestação de serviços. Ele registra, no entanto, sua forte impressão quanto à importância social adquirida pela codificação do conhecimento teórico e a centralidade que ele conquistou na geração tanto de novos conhecimentos quanto de bens econômicos e serviços.

Com relação à importância que atualmente é dada à informação, Kumar (1997) afirma que a explosão informacional não produziu uma mudança radical no modo como as sociedades industriais são organizadas, ou, na direção para onde vinham se movendo.

“Os imperativos de lucro, poder e controle são tão predominantes agora quanto foram na história do capitalismo industrial. A diferença está na variedade e na intensidade de suas aplicações... não em alguma mudança nos próprios princípios”. (KUMAR, 1997, p. 191).

A 'sociedade da informação' não avançou para seu atual patamar através transformações quantitativas, mas, qualitativas. Herbert Schiller (1981), com apurada percepção, afirma que não foi a quantidade de informação que alterou nosso modo de vida, mas, o que fomos capazes de fazer com essa informação.

Transformar o modo de vida implica um comportamento *inteligente*, o qual envolve conhecimento e não simplesmente informação ou tecnologia. A mudança qualitativa observada na dinâmica social (da chamada sociedade da informação) é a transformação qualitativa de mera informação em conhecimento. A ciência, nesse contexto, aparece no centro da sociedade da informação impulsionando o desenvolvimento do conhecimento.

É difícil argumentar contra o fato de que o conhecimento ocupa uma posição preeminente na sociedade atualmente, contudo, o próprio Giddens (1985, 2002) é um estudioso que permanece relutante em reconhecer nessa preeminência um aspecto histórico inédito. Esse fenômeno não parece ser tanto uma revolução, mas, um processo que agora alcançou uma etapa avançada de seu desenvolvimento.

O conhecimento, não é difícil ver, sempre ocupou uma posição central nos assuntos humanos. Sua importância se estabeleceu de forma inelutável nas mentes de sucessivas gerações que experimentaram seu efeito e sua eficácia nas mais diferentes atividades do cotidiano.

Talvez, o epíteto sociedade da informação se deva, de fato, à atenção que hoje é dada à codificação do conhecimento. É sensível o modo como a sociedade se mostra comprometida com a formalização e o registro do conhecimento (a informação, como proposta por Foskett (1998)) e de diversos tipos de experiência que podem inculcar algum tipo de conhecimento. Briggs e Burke (2006) observam que:

“[...] a nova expressão 'sociedade da informação' dava forma ou moldava um conjunto, até agora organizado de forma frouxa, de aspectos relacionados à comunicação – conhecimento, notícias, literatura, entretenimento – todos permutados entre mídias e elementos de mídias diferentes – papel, tinta, telas, pinturas, celulóide, cinema, rádio, televisão e computadores. Da década de

1960 em diante, todas as mensagens, públicas ou privadas, verbais ou visuais, começaram a ser consideradas ‘dados’, informação que podia ser transmitida, coletada e registrada, qualquer que fosse seu lugar de origem, de preferência por meio de tecnologia eletrônica.” (BRIGGS e BURKE, 2006, p. 260).

Manter registros tornou-se uma parte fundamental da maioria das atividades modernas. Esses registros vão compor os repositórios de conhecimento que possibilitam a reflexão social e a auto-reflexão que caracterizam nosso modo de vida (GIDDENS, 2002). Esses registros constituem o complexo ambiente informacional em que vivemos.

No que diz respeito à ciência, o registro do conhecimento produzido segundo seus métodos – em anais de encontros científicos, artigos, relatórios técnicos e patentes – consolidou-se de forma inapelável como fundamento do trabalho científico. O modelo atual da comunicação científica conseguiu, através do periódico científico, estabelecer definitivamente o conhecimento teórico como ferramenta de transformação da realidade social.

## **2.2 A comunicação científica**

Podemos afirmar que o contato entre homens de ciência sempre existiu (WEISMAN, 1972), (ROSSI, 1992). O estudo da comunicação científica, no entanto, é um fenômeno recente. Esse termo foi cunhado por John Desmond Bernal no final da década de 1930 para designar “o processo específico de produção, consumo e transferência da informação no campo científico” (CUNHA e CAVALCANTI, 2008, p. 97).

Segundo Mueller (2007), são três seus iniciadores: Paul Otlet, Vannevar Bush e Alexander Ivanovich Mikhailov. Mueller (2007) afirma que o estudo da comunicação científica, por meio de sua literatura, está intimamente ligado às origens da própria ciência da informação.

Otlet divisou e propôs formas de controlar a informação para garantir o acesso a ela. Bush traçou como seu objetivo dar encaminhamento ao problema do armazenamento e do acesso à informação. Mikhailov inaugurou uma nova disciplina para o estudo da estrutura e das propriedades gerais da informação científica.

Mikhailov *et al.* (1980) defendem que o objeto da 'informática' é bem definido:

"visto que a expressão anglo-americana ciência da informação pode levar alguém a interpretá-la num sentido amplo, isto é, como uma ciência que trata de todos os tipos de informação, da informação em geral, julgamos necessário enfatizar que a palavra informação aqui significa apenas informação científica." (MIKHAILOV *et al.*, 1980, p. 72).

A comunicação científica pode ser entendida como o processo de construção do conhecimento pela comunidade científica. Na verdade, esse processo dá "acabamento" ao conhecimento construído. Através da comunicação científica é possível a confirmação e validação dos novos conhecimentos em um nível plural de problematização, de modo que esses novos conhecimentos se tornem parte de nossa experiência coletiva. Conforme Ziman (1979) explica:

"Quando dizemos que o conhecimento científico é empírico, ou grandemente baseado em provas empíricas, isso não quer dizer que todo cientista tenha visto com seus próprios olhos todas as maravilhas em que acredita. O que estamos pretendendo dizer é que ele pode dispor de uma série de informes sobre pesquisas feitos por testemunhas dignas de crédito e apresentados de acordo com certas convenções, de maneira clara e sem ambiguidades, além de ser dada a devida atenção a possíveis fontes de erro. Esses informes não são um diário que se proponha a nos contar exatamente o que se passou num determinado dia em determinado laboratório. [...] [mas,] ao se tornarem parte do acervo de conhecimento público, eles se transformam em provas "de ouvido", em informações de segunda-

mão não inteiramente distantes da experiência direta realizada por um de nós.” (ZIMAN, 1979, p. 50).

O impacto que a codificação do conhecimento tem na sociedade da informação é percebido claramente nessa explanação de Ziman (1979).

Silva *et al.* (2006, p. 172) argumentam que a produção científica realmente não se limita à realização de pesquisa, enquanto puro processo investigativo, mas, envolve de forma equivalente a comunicação dos resultados desta pesquisa. “O processo de comunicação é de suma importância, pois a divulgação dos resultados obtidos é o primeiro passo para que o conhecimento científico adquira confiabilidade e os autores credibilidade e prestígio.”

O interesse pela comunicação científica como disciplina começa no século XX e foi correspondente ao estímulo que o provocou: o crescimento da literatura científica. Contudo, o crescimento desse tipo de literatura não é um fenômeno circunscrito ao século XX. Com a vulgarização do uso da imprensa de tipos móveis, a quantidade de informações em fluxo no meio filosófico e científico se tornou motivo de preocupação e reflexão.

“A própria publicação de livros era um negócio que atraía o interesse de negociantes que já ajudavam a financiar impressores no século XV. Ainda mais importante [...] era o fato de que a impressão encorajava a comercialização de todos os tipos de conhecimento. Uma consequência óbvia, mas significativa, da invenção da imprensa foi envolver os empreendedores de maneira mais direta no processo de difusão do conhecimento, “o negócio do iluminismo” [título de um livro de Robert Darnton]. Os impressores às vezes encomendavam novas edições de textos clássicos, traduções e obras de referência.

A frequente publicação quase simultânea de obras diferentes sobre o mesmo assunto dá uma ideia da intensidade da concorrência entre impressores, da mesma forma que a reivindicação, nas folhas de rosto, de que a nova edição era mais acurada que as anteriores, ou

incluía mais informação, ou que oferecia um sumário ou um índice que faltava às outras.” (BURKE, 2003, p. 145-146).

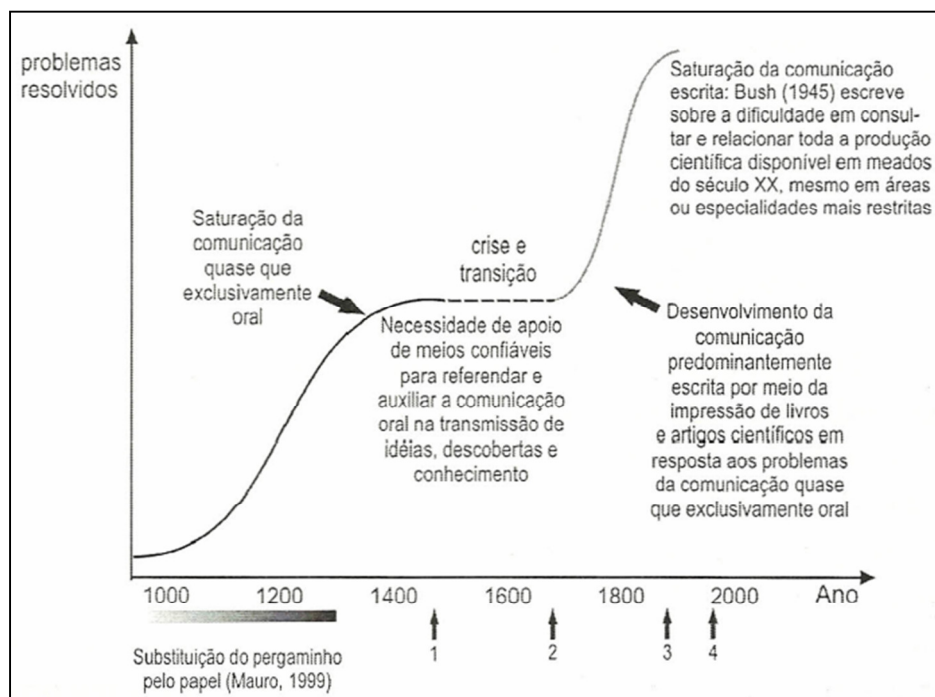
O número de livros publicados desde então até o século XVII cresceu muito e chegou a ser tratado como “uma das doenças desta época” que “sobrecarregam o mundo” (PRICE, 1976, p. 40).

“O aparecimento do periódico científico trouxe a esperança de se por um fim a essa sobrecarga iníqua. Desenvolvendo-se com o tempo e em espírito juntamente com o jornal, publicações como *Philosophical Transactions* da Sociedade Real inglesa, tinham a função definida de resumir os livros e fatos da cultura em toda a Europa.” (PRICE, 1976, p. 40).

Aos poucos, o formato do novo periódico científico foi sendo adotado em grandes centros por toda a Europa, difundindo-se a partir da França e do Reino Unido onde foi primeiro concebido. À medida que crescia, ainda mais, o número de volumes impressos (livros e periódicos) circulando pelo velho mundo, cresciam também as pressões para que os intelectuais fizessem contribuições relevantes ao conhecimento (BURKE, 2003) (ZIMAN, 1979), fenômeno que dinamiza ainda mais o processo de produção de informações.

O periódico científico não solucionou efetivamente o problema da sobrecarga da informação, mas, permitiu que os resultados das pesquisas científicas conseguissem uma ampla exposição e de uma maneira eficiente. Seu uso acabou por redesenhar a própria feição da comunidade científica, estabelecendo um novo modelo de interação entre os pesquisadores.

Côrtes (2006) entende a história da comunicação científica de forma dialética. Para ele, a evolução no processo de comunicação científica ocorreu de forma paradigmática conforme o GRÁF. 1 abaixo.



**GRÁFICO 1 – A evolução dos paradigmas da comunicação científica**

Fonte: CÔRTEZ, 2006, p. 52.

Nesse diagrama, 1) corresponde aos avanços na impressão de Gutenberg; 2) corresponde ao nascimento do periódico científico; 3) corresponde à transformação dos artigos científicos para o seu formato atual; e, 4) corresponde ao momento de saturação do modelo de comunicação vigente. Esse diagrama ilustra bem o salto qualitativo que a adoção dos periódicos científicos proporcionou ao processo de comunicação científica.

Para Côrtes (2006), as inovações tecnológicas associadas ao processo de comunicação científica estão permitindo que a difusão do conhecimento científico possa ascender a um novo paradigma. Esse novo paradigma traz, implicado a ele, soluções adequadas aos problemas germinados desde a instauração do paradigma anterior, onde a comunicação científica se processava exclusivamente por canais impressos. A situação atual do modelo de comunicação científica é caracterizada por Côrtes (2006) como de “saturação do modelo” e as novas tecnologias participam formidavelmente na mudança desse estado de coisas.

Weitzel (2006) sustenta que os estudos sobre documentação científica e comunicação científica podem ser entendidos de acordo com a ênfase que adotam. Estudiosos que se debruçaram sobre esse tema tenderam a enfatizar algum dos processos envolvidos: seja geração, disseminação, ou, uso da informação.

Para Weitzel (2006), Francis Bacon (1952), Robert Boyle (1999), e Robert Merton (1979) discutem aspectos do acervo produzido pela ciência, colocando ênfase na geração do conhecimento. John Bernal (1954, 1976) e William D. Garvey (1972) colocam ênfase na sua disseminação. Paul Ginsparg (1997) e Stevan Harnad (1990, 2004) colocam ênfase no uso do conhecimento.

A ênfase na geração do conhecimento está ligada a considerações sobre a instituição do *sistema de informação da ciência* e de sua estrutura social.

“A ênfase na disseminação do conhecimento científico emerge da necessidade de aperfeiçoar os mecanismos de disseminação do conhecimento em função da explosão bibliográfica e fragmentação do conhecimento científico. Bernal é considerado o arauto e Garvey com sua equipe procuraram mapear o fluxo da informação para promover a inovação desse sistema que já apresentava sinais de esgotamento na ocasião.” (WEITZEL, 2006, p. 89).

A ênfase na otimização do uso do conhecimento surge com as tensões entre os interesses comerciais de grandes casas editoriais, proprietárias de títulos de periódicos científicos, e os interesses acadêmicos dos pesquisadores autores de artigos.

Sob qualquer ênfase, e tanto no paradigma impresso quanto no paradigma digital, os estudos de comunicação científica revelam o caráter superlativo do periódico científico no que pode ser entendido como um processo de “estabilização das redes de conhecimento” (FROHMANN, 1999) entre pesquisadores. As redes que põem em contato os pesquisadores são fundamentais para a definição de comunidade científica e para a definição de teoria científica.

Frohmann (1999) afirma que a função prática que o periódico científico cumpre não é propriamente a de geração de conhecimento, mas, a de consolidação, de estabilização de um determinado modelo teórico compartilhado.

Para Frohmann (1999) da mesma forma que para Ziman (1979), os artigos científicos não são uma expressão exata do processo de descoberta científica. Seu propósito basilar é a unificação da ciência. Para Frohmann (1999) tal unificação consiste no esforço de cooptar a comunidade científica em torno de um discurso único; para Ziman (1979) essa unificação é alcançada pela própria dinâmica social da comunidade científica quando um determinado discurso produzido em seu meio é aceito como conhecimento público.

O conteúdo mais repercutido dos artigos científicos em uma determinada área do conhecimento é o que dá ao seu campo conceitual e teórico um caráter homogêneo. “A escrita formal”, segundo Frohmann, “é crucial para o estabelecimento das técnicas documentárias para a enunciação institucionalmente autorizada da verdade científica.” (FROHMANN, 1999, p. 72).

Segundo Campello e Campos (1993), Mueller (1994, 1999), Miranda (1996), e Ferreira *et al.* (2004), as principais funções da comunicação científica são: registro de contribuição científica; certificação do conhecimento científico; disseminação da informação; e memória.

Herbert Menzel, em um manuscrito de 1958, intitulado: “*The Flow of Information among Scientists: Problems, Opportunities and Research Questions*” (KAPLAN e STORER, 1968) afirma que o processo de comunicação científica:

- Fornece respostas para perguntas específicas;
- Ajuda o cientista a se manter informado e a entender um novo campo de interesse;
- Expõe as principais tendências de uma área e dá ao cientista uma noção da importância do seu trabalho;
- Verifica a confiabilidade de uma informação pela adição de testemunhos;
- Permite, ao cientista, uma auto-avaliação de seus interesses;

- Permite que o cientista acesse respostas críticas ao seu trabalho.

Dentre todas essas funções destaca-se a da contribuição científica, que se tornou um dos critérios mais importantes para a promoção da carreira de um acadêmico e de um pesquisador (MEADOWS, 1999), (ZIMAN, 1979), (CAMPELLO e CAMPOS, 1993), (MUELLER, 1994, 1999).

Frente a um fazer científico marcado por vigorosa produção escrita, as facilidades introduzidas pela tecnologia vieram permitir que novos canais de comunicação e informação fossem experimentados como formas alternativas para o escoamento dessa produção escrita. A importância do periódico científico se redefiniu de forma especial nesse contexto (como veículo formal já estabelecido), pois, os novos canais criados também não se mostraram capazes de resolver pontos vulneráveis do processo de comunicação científica, como o próprio estabelecimento formal da contribuição científica de um autor (MUELLER, 1994).

Os canais de comunicação caracterizam o processo de comunicação e seu fluxo. O fluxo da comunicação científica foi representado com grande êxito por William Garvey e Belver Griffith, em 1972. Eles criaram o modelo da FIG. 1 que mostra como autor e comunidade científica se relacionam com o artigo científico no intuito de incorporar seu conteúdo ao estoque de conhecimento.

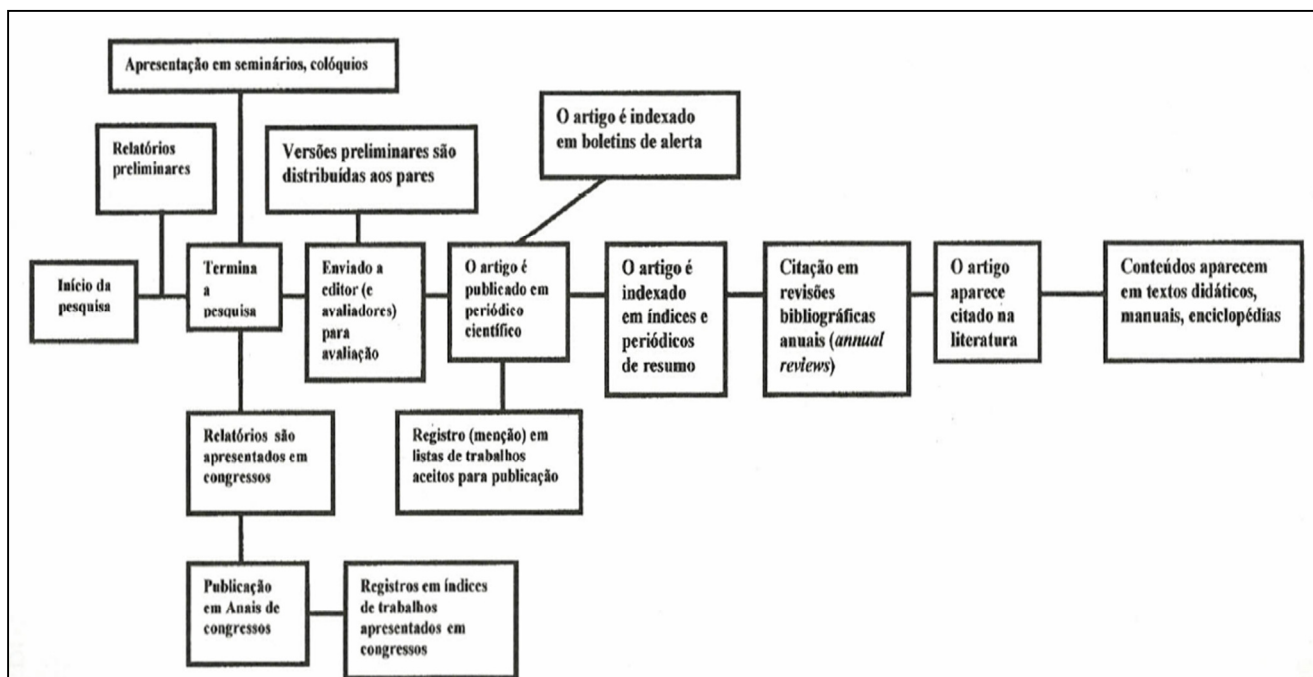


FIGURA 1 – Modelo de comunicação científica de Garvey e Griffith

Fonte: MUELLER, 2007, p. 29.

Nesse modelo as etapas anteriores à publicação do artigo científico constituem o que chamamos de canais informais da comunicação científica. As etapas posteriores são chamadas de canais formais.

Nos canais informais, o conteúdo veiculado ainda está muito ligado ao autor. A informação é recém-produzida e o autor ainda sente necessidade de fazer ajustes na forma de sua apresentação; a audiência é relativamente pequena; a informação é de difícil recuperação. Nos canais formais, a informação já se apresenta bem elaborada e as conceitualizações bem formadas; a audiência é ampla; o processo de recuperação da informação dispõe de mecanismos específicos; e a preservação do conteúdo é prevista e arranjada.

Segundo Población e Oliveira (2006) na fase informal da comunicação científica é gerada a literatura cinzenta e os resultados da pesquisa são consolidados e evoluem para a fase formal. Na fase formal, não só autores, mas editores e centros de informação são responsáveis pela qualidade do conteúdo e por aspectos formais dos documentos publicados. Para aquelas autoras, é a fase formal que apresenta os

maiores desafios relacionados à disseminação, visibilidade e acessibilidade da informação.

“Os sucessos, ou insucessos, dessa etapa decorrem das estratégias e dos mecanismos encontrados para difundir a pesquisa nos veículos adequados, garantir a visibilidade nas bases de dados e facilitar o acesso não só ao resumo, mas também, ao texto completo.”  
(POBLACIÓN e OLIVEIRA, 2006, p. 72).

As facilidades tecnológicas introduzidas nos últimos anos quase desfizeram a distinção entre canais formais e informais, transfigurando todo o fluxo da comunicação científica. No cenário atual de dura competitividade e de grande produção acadêmica, os fluxos de comunicação ligados às redes sociais de caráter informal são utilizados com muito interesse por pesquisadores em geral. Esses fluxos apresentam características de um canal formal: um leiaute sofisticado – dando à forma discursiva uma configuração mais elaborada; o número de pessoas atingidas é enorme; e, a recuperação da informação é relativamente fácil. Nessas circunstâncias a distinção entre formal e informal se baseia unicamente no sistema de avaliação pelos pares.

A estrutura essencial de um periódico científico envolve um editor, um corpo editorial, e um arranjo administrativo. Na avaliação pelos pares, um trabalho submetido para publicação em determinado periódico é enviado para dois especialistas membros do corpo editorial para revisão e indicação do trabalho para publicação no periódico. Esse processo valida o conhecimento e autoriza sua divulgação e uso.

O sistema de avaliação pelos pares – embora consagrado a nível internacional – imprime alguns traços indesejáveis ao fluxo de comunicação. Alguns deles são: perpetuação de padrões na distribuição de recursos; determinação de prioridades de pesquisa; e, baixa eficiência na reestruturação e redirecionamento das atividades científicas (MENEHINI e FONSECA, 1990). Esses traços se tornam indesejáveis principalmente porque influenciam o fluxo da informação a partir de preceitos não puramente intelectuais ou de interesse exclusivamente científico.

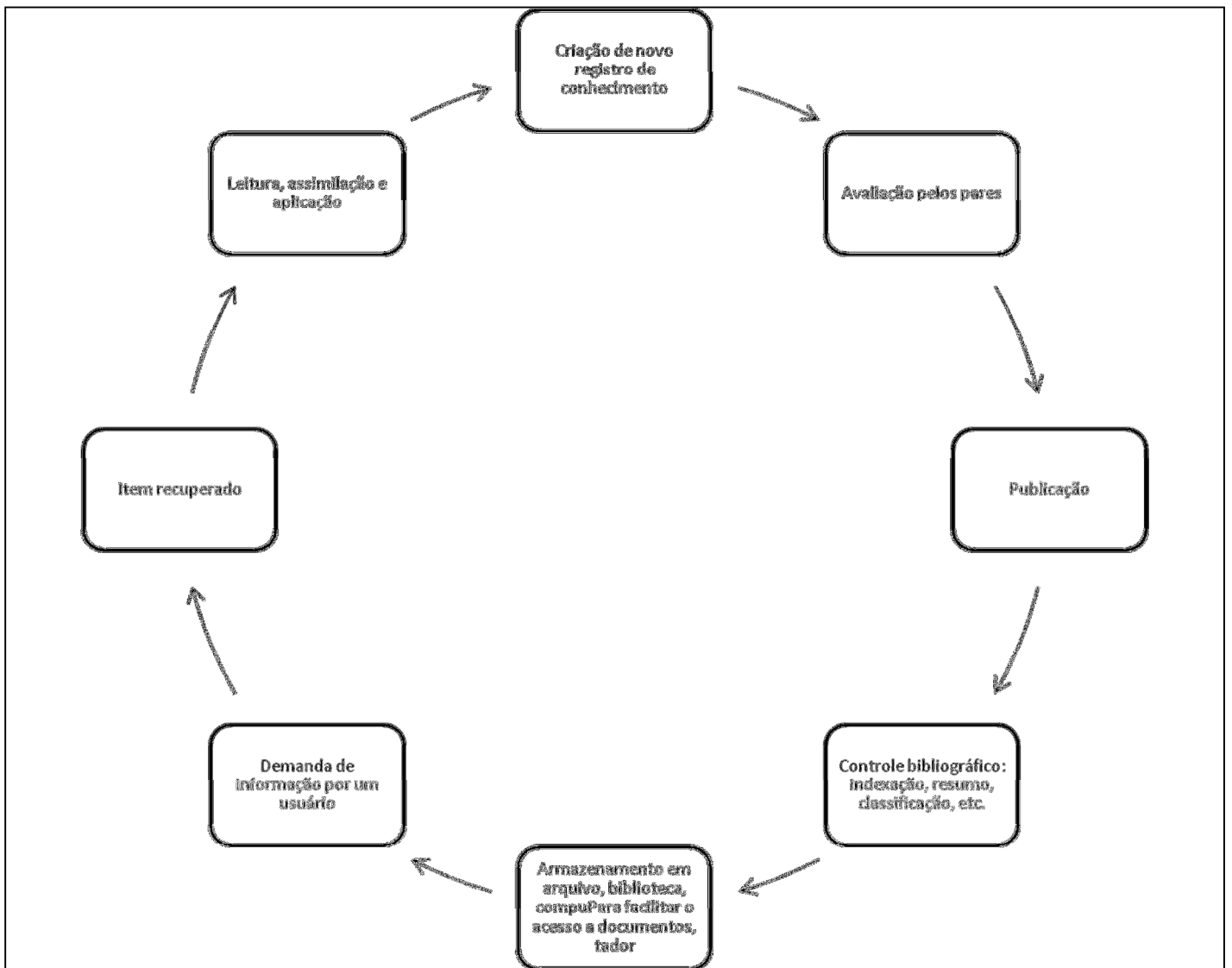
O aumento do fluxo de comunicação científica é marcado não só pelo aumento no número de artigos científicos, mas, também, pela proliferação dos periódicos científicos. Esse é um fenômeno característico dos canais formais.

Conforme Mueller (2000), essa proliferação dos periódicos reflete em parte o crescimento normal da ciência, mas, as regras da própria comunidade científica também têm contribuído artificialmente para isso. Os sistemas de promoção na carreira, concessão de prêmios, e, financiamentos adotam o volume de publicações de um pesquisador (que é estipulado como sua contribuição científica) como um dos critérios mais importantes no julgamento do mérito científico.

As próprias regras da comunidade científica, então, têm concorrido para o desgaste da noção de contribuição científica, para a lentidão no processo de avaliação pelos pares e para a dispersão da literatura. Na maioria dos casos, um pesquisador aceitará publicar seus trabalhos mesmo em periódicos científicos que não sejam da área da sua especialidade. Essa experiência concorre para a interdisciplinaridade, é verdade, mas da forma como é implementada, contribui também para tornar a informação menos visível e menos acessível, tornando essa interdisciplinaridade clandestina.

As consequências das pressões que tornam miscíveis canais formais e informais; da determinação “política” de prioridades de pesquisa; dos sistemas vigentes de promoção na carreira – que ajudam a corroer o conceito de contribuição científica; as consequências de tudo isso são: disseminação e recuperação ineficientes da informação científica.

O processo de formação de estoques de conhecimento é característica do chamado ‘ciclo informacional’ (FIG. 2). As atividades envolvidas no ciclo informacional – que transformam informação em conhecimento – vêm atendendo, de alguma forma, as necessidades de informação da sociedade, mas, precisam avançar no que diz respeito ao refinamento de suas relações recíprocas e à qualidade dos resultados obtidos em cada etapa do ciclo.



**FIGURA 2 – Ciclo Informacional**

Fonte: CLEVELAND e CLEVELAND, 2001, p. 23.

Na FIG. 2 a disseminação da informação é representada pelo semicírculo direito e a recuperação da informação, pelo semicírculo esquerdo. As etapas de: avaliação pelos pares; controle bibliográfico; demanda de informação (na verdade, o atendimento a essa demanda); e, leitura, assimilação e aplicação constituem gargalos no ciclo informacional.

O que a atividade de avaliação pelos pares representa para a validação e publicação do conhecimento científico, as atividades de indexação e elaboração de resumos

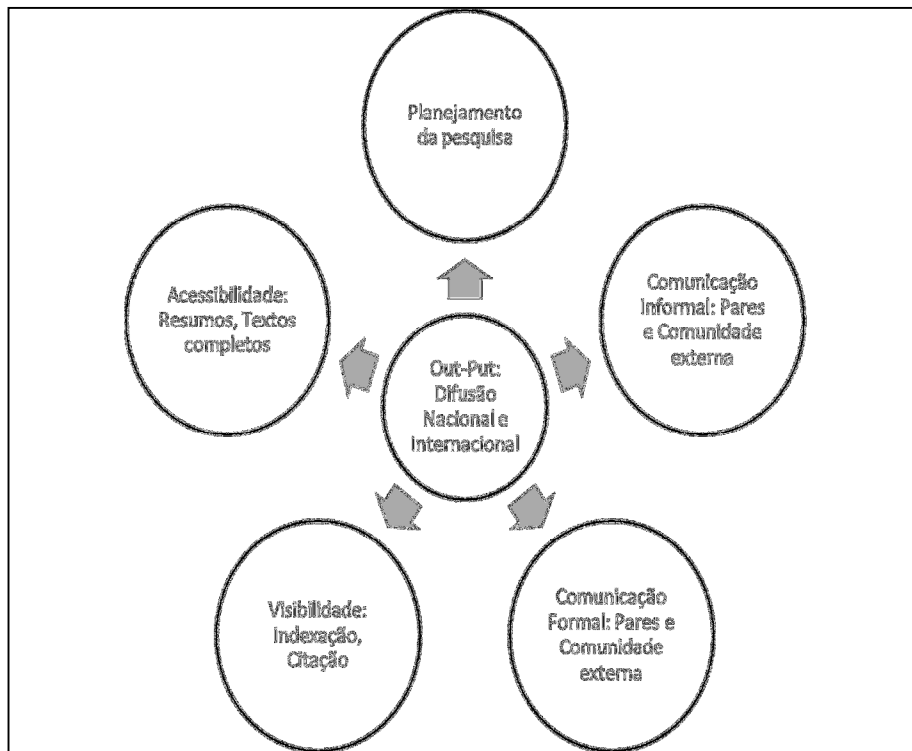
representam para o controle bibliográfico e para o atendimento à demanda por informação.

A etapa definitiva do ciclo informacional, e que define a sociedade da informação, é a geração de nova informação e de novo conhecimento. O profissional da informação participa da geração de novo conhecimento de modo categórico: ele participa de sua colocação a nível social. Seu ofício é o de aproximar certos tipos de conhecimento discriminando-os de outros tipos de conhecimento. Essa determinação tem efeito público, inclusive sobre a noção de relevância da informação. Seu ofício é a identificação, a descrição e a classificação do conhecimento, definindo sua abrangência e alcance, ou, sua especialização e acuidade.

Os índices e resumos produzidos pelo profissional da informação estão relacionados à visibilidade e acessibilidade na comunicação científica. É trivial a noção de que índices e resumos incrementam a visibilidade e a acessibilidade de conteúdos informacionais tratados, entretanto, a essa noção gostaríamos de acrescentar a distinção explorada nos parágrafos que se seguem.

Na FIG. 3, o efeito mais notável da atribuição de descritores a um determinado documento é aumentar a *visibilidade* desse documento. Isso acontece porque esse documento é apresentado sob a perspectiva panorâmica do conceito (a que o descritor se refere), o qual representa não só esse documento mas também outros que tratam do mesmo assunto. A ampliação da visibilidade desse documento hipotético ocorre junto com a ampliação da visibilidade dos outros documentos mencionados porque eles são conceitualmente análogos.

Os descritores de um índice oferecem *indícios* a respeito dos documentos que descreve e, meramente, situam o usuário no ambiente informacional no qual se insere. Meramente, relacionam o usuário a um subconjunto de documentos de potencial interesse.



**FIGURA 3 – Visibilidade e Acessibilidade da Produção Científica**

Fonte: Adaptado de POBLACIÓN e OLIVEIRA, 2006, p. 71.

O resumo, por sua vez, aumenta a *acessibilidade* de um documento. O resumo aumenta a acessibilidade de um documento porque oferece ao usuário a garantia de que sua consulta ao texto original integral será recompensada (ainda que isso aconteça, na experiência concreta, com algumas tentativas frustradas devido a resumos mal elaborados).

O resumo oferece *evidência* que direciona o usuário no ambiente informacional, relacionando-o univocamente a um documento de efetivo interesse. Entretanto, a acessibilidade que o resumo proporciona só é segura no caso de resumos informativos, resumos indicativos não possuem essa propriedade (veja a discussão feita nas pág. 63-64, na seção “A elaboração de resumos”).

O uso de resumos em comunicação científica foi consagrado pela comunidade de praticantes devido ao seu potencial no processo de disseminação e acesso da informação. A dupla função do resumo, informativa e de representação, o caracteriza

como um elemento próprio para a circulação e recuperação da informação, aumentando a eficiência desse processo.

Price (1976) e Herschman (1970) observaram, há muito tempo, que dentro de um processo comunicativo formal a função de disseminação da informação torna-se cada vez mais dependente da função de recuperação da informação.

## **2.3 A recuperação da informação científica**

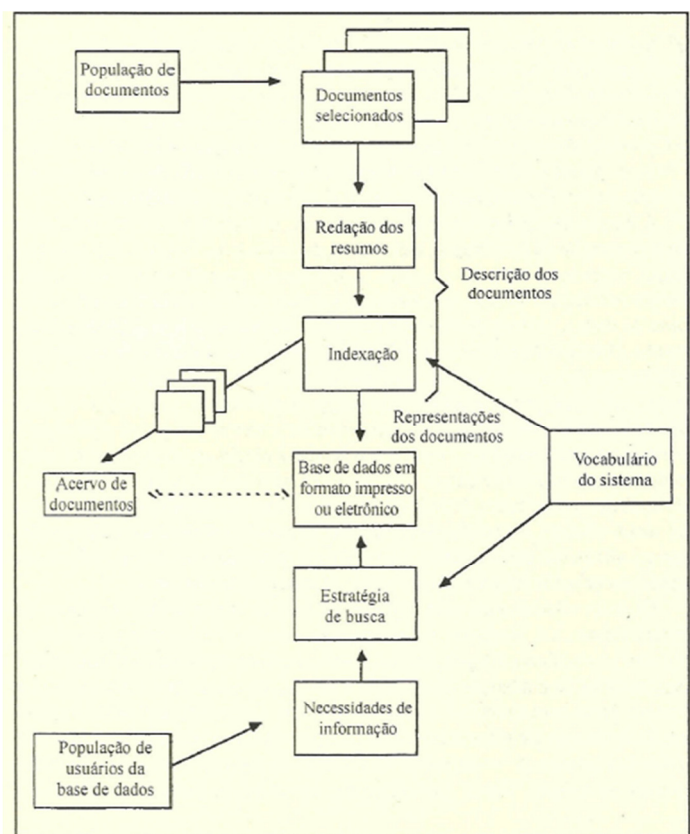
Os estoques de conhecimento e informação são a pedra angular de um sistema de comunicação formal. De sua organização intelectual depende a eficiência de todo o processo comunicativo. No processo de comunicação científica isso não é menos verdadeiro.

Os documentos em bibliotecas tradicionais, por exemplo, são organizados por assuntos abrangentes e suas derivações. A organização desses assuntos evidencia, em algum grau, relações entre os próprios documentos. Além disso, o catálogo preparado pela biblioteca dá ao usuário um referencial tornando a atividade de recuperação da informação uma atividade cognitiva com clareza estratégica. Em um contexto como este, a organização da informação é planejada de forma que a inércia da recuperação da informação seja mínima.

Na Internet o acesso remoto e imediato a documentos aparece como o grande diferencial em relação aos acervos tradicionais. Infelizmente, os documentos na rede são estruturados precariamente. O usuário procede em sua busca autônoma, mas lenta e erraticamente sem um referencial e sem uma relação significativa evidenciada entre os documentos.

Na FIG. 4 a recuperação da informação é retratada no contexto das atividades rotineiras de um estoque de conhecimento (ou uma base de dados) qualquer. O produtor de uma base de dados seleciona documentos para a inclusão em seu acervo

segundo critérios definidos. Os documentos selecionados são descritos de acordo com os processos de tratamento da informação adotados naquela base de dados. Os termos utilizados para a descrição dos documentos formam o 'vocabulário do sistema' que servirá como interface entre a base de dados e seus usuários. Os usuários com uma necessidade de informação se dirigem à base de dados e elaboram alguma estratégia de busca para consultar os documentos da base de dados.



**FIGURA 4 – A recuperação da informação**

Fonte: LANCASTER, 2004, p. 2.

No que diz respeito à ação do usuário representada na FIG. 4, essa ação se processa da mesma maneira na recuperação da informação científica.

O processo de recuperação da informação é um processo que envolve conhecimento. A interação do usuário com uma base de dados pode ser vista como uma atividade cognitiva ou um sistema cognitivo global definido da seguinte forma:

“Questão 1: O que é a cognição?”

Resposta: É o tratamento da informação: a manipulação de símbolos a partir de regras.

Questão 2: Como funciona?

Resposta: Por meio de qualquer dispositivo que possa representar e manipular elementos físicos descontínuos: os símbolos. O sistema interage apenas com a forma dos símbolos (os seus atributos físicos) e não com o seu sentido.

Questão 3: Como saber se um sistema cognitivo funciona de forma apropriada?

Resposta: Quando os símbolos representam adequadamente qualquer aspecto do mundo real, e o tratamento da informação propõe uma solução eficaz para o problema submetido ao sistema.” (VARELA, 199?, p. 34).

Está claro que o acesso eficiente à informação superabundante acumulada nos estoques de conhecimento depende essencialmente do tratamento que essa informação recebe.

Segundo Dias (2001):

“Nos sistemas de informação e de recuperação da informação, o tratamento da informação é definido como a função de descrever os documentos, tanto do ponto de vista físico (características físicas dos documentos) quanto do ponto de vista temático (ou de descrição do conteúdo). Essa atividade resulta na produção de representações documentais (fichas de catálogo, referências bibliográficas, resumos, termos de indexação, etc.) que não apenas se constituem de unidades mais fáceis de manipular num sistema de recuperação da informação (comparado ao documento em sua íntegra), como também representam sínteses que tornam mais fácil a avaliação do

usuário quanto à relevância que o documento integral possa ter para as suas necessidades de informação.” (DIAS, 2001, texto em html).

No processo de recuperação da informação – como evidenciado pela FIG. 4 – as atividades de indexação e resumo são fundamentais. Ambas as atividades implicam a preparação de uma representação do conteúdo de documentos. Na indexação essa representação é feita por meio de uma lista de termos ou expressões que procuram transmitir o teor do documento que representam. No resumo, essa representação é feita por meio de um pequeno texto que procura transmitir a singularidade de um determinado documento.

Tanto na indexação quanto na elaboração de resumos cada documento é tratado em duas etapas: análise conceitual e tradução. Na análise conceitual, o conteúdo é interpretado segundo determinados parâmetros. São identificados os conceitos que definem a extensão temática do documento. Esses conceitos são cotejados com outros, internamente pelo indexador, a fim de se determinar sua especificidade e sua adequação em relação ao documento. Na etapa da tradução, cada um desses conceitos é relacionado com o termo do vernáculo ou do vocabulário do sistema que melhor denota seu significado e esse termo se torna um termo descritor desse documento.

Os descritores arrolados em um índice – organizado para a recuperação dos documentos de um acervo – constituem-se em apontadores para documentos, *indicando* seu conteúdo. Um grande número de termos descritores atribuídos a um documento aumenta a possibilidade da recuperação desse documento em buscas bastante diversas. Se os documentos de um acervo forem assim tratados, as buscas nesse acervo terão uma alta ‘revocação’. Isso quer dizer que serão produzidos abundantes resultados relevantes para cada busca, mas, junto com eles, uma enormidade de itens irrelevantes.

Por outro lado, um número pequeno de termos descritores – coerente e especificamente atribuídos a um documento – aumenta a possibilidade da recuperação desse documento *somente* em buscas caracterizadas por uma

proximidade temática. Se os documentos de um acervo forem assim tratados, as buscas nesse acervo terão uma alta 'precisão'. Isso quer dizer que serão produzidos resultados em pequeno número, mas, resultados compensadores. No entanto, várias buscas serão necessárias para que seja recuperado o maior número possível de documentos relevantes (existentes nesse acervo) para o assunto.

Podemos afirmar, comparando índices e resumos, que o índice exerce uma influência na revocação de resultados, e, o resumo, na precisão deles.

É oportuno aqui discutir alguns aspectos teóricos relacionados à recuperação da informação.

Se conhecimento é crença verdadeira justificada – conforme a lição de Platão em Teeteto – nosso interesse pelo conhecimento registrado em um lugar qualquer é da medida de nossa adesão às proposições que justificam esse conhecimento. É de se notar a pertinência das críticas levantadas a essa definição de conhecimento, mas mesmo com todas as dificuldades que ela suscita, ela ainda apela fortemente à nossa intuição e oferece uma fórmula sugestiva como modelo de conceituação e reflexão.

Defende-se aqui que o tratamento da informação precisa observar as implicações epistemológicas e os desdobramentos teóricos de cada registro de conhecimento. O tratamento da informação precisa se estruturar formal e sistematicamente em torno desse eixo conceitual. A preeminência do conhecimento teórico na sociedade da informação parece prescrever o processamento da informação nesses termos.

A busca por informação é orientada por uma necessidade de informação volátil, mas, animada por pressupostos teóricos definidos, ainda que implícitos. O indivíduo alimenta suas expectativas na busca com base em seus pressupostos e em suas crenças (CARNEIRO, 1982); (CRAWFORD, 1978). Desse modo, a informação precisa ser tratada de modo a contemplar crenças e pressupostos subjacentes às necessidades informacionais. O tratamento da informação deve identificar nos documentos em geral os padrões que estruturam as crenças e os pressupostos implicados, os modelos teóricos responsáveis pela coesão do seu conteúdo informativo.

Ferreira (1997, p. 5) discute problemas nos estudos de necessidade e uso da informação e constata que:

- falta uniformidade conceitual às pesquisas – ‘informação’, ‘necessidade de informação’ e ‘uso de informação’ são termos usados indiscriminadamente (e nesse aspecto os pesquisadores reproduzem os equívocos dos usuários da informação);
- faltam definições e pressupostos claros para focalizar variáveis e gerar questões de pesquisa;
- não há metodologias específicas e com rigor científico.

Os profissionais da informação envolvidos com a comunicação científica são responsáveis para que avanços na comunicação do conhecimento ocorram. Conforme Wersig e Neveling (1975, p. 129): "transmitir o conhecimento para aqueles que dele necessitam é uma responsabilidade social, e essa responsabilidade social parece ser o verdadeiro fundamento da ciência da informação".

O espaço interdisciplinar, e mais do que isso, o espaço epistemológico das ciências parece o lugar próprio para esse trabalho. O profissional da informação se encontra no espaço “interdocumental” e da mesma forma que procura por relação entre diferentes documentos (de conteúdo disciplinar) pode encontrar relações entre diferentes disciplinas. Ele precisa, no entanto, aprofundar seu conhecimento a respeito de diferentes perspectivas teóricas de vários campos do conhecimento humano a fim de poder contribuir efetivamente para a “liquidez” do conhecimento entre as diferentes disciplinas.

Leonard (1975) fez um estudo sobre consistência entre indexadores e eficiência na recuperação da informação. Ele não encontrou uma correlação positiva entre coerência na indexação e formação educacional. Segundo os dados encontrados, não é de se esperar que, mesmo indexadores especialistas no tema tratado por um determinado documento, concordem em sua indexação desse documento (LANCASTER, 2004, p. 76).

Esse achado nos ajuda, enquanto profissionais da informação, a desfazer o sentimento de acanhamento e imobilizante prudência em relação a uma área do conhecimento na qual não somos especialistas.

Mesmo ciências consolidadas como a Física, por exemplo, precisam avançar em aspectos de sua discussão epistemológica. Os professores Antônio Máximo e Beatriz Alvarenga (1997) assim introduzem o seu *Curso de Física*:

“A Física, no início de seu desenvolvimento, era considerada como a ciência que se dedicava a estudar todos os fenômenos que ocorrem na natureza. Daí ter sido esta ciência, durante muitos anos, denominada “Filosofia Natural”.

Entretanto, a partir do século XIX, a Física restringiu seu campo, limitando-se a estudar mais profundamente um menor número de fenômenos, denominados “fenômenos físicos” e os fenômenos que dela se destacaram deram origem a outras ciências naturais.

Se tentássemos, porém, esclarecer quais são os chamados “fenômenos físicos”, verificaríamos que não seríamos capazes de estabelecer uma definição clara. Mas não nos preocupemos com isto. Com o desenrolar deste curso, você irá descobrindo que é mais importante saber e compreender o que já se fez no campo da Física, mesmo que não se possa defini-la em poucas palavras.” (MÁXIMO e ALVARENGA, 1997, p. 5).

No tratamento da informação, as tarefas de análise conceitual e tradução são em tudo semelhantes a um exercício epistemológico, só que se trata de uma epistemologia aplicada.

Segundo Dancy (1995):

“Epistemologia é o estudo do nosso direito às crenças que temos. De forma mais geral, nós começamos com o que podemos chamar de atitudes cognitivas, e então, perguntamos se fazemos bem em ter tais atitudes. Atitudes cognitivas incluem tanto nossas crenças (o que

nós tomamos ser) quanto nosso saber; e em outra dimensão, elas incluem nossa postura com relação às várias *estratégias e métodos* que usamos para conseguir novas crenças e filtrar as antigas, bem como, *os produtos* dessas estratégias e métodos.” (DANCY, 1995, p. 245, o grifo é nosso).

Na atividade documentária, as tarefas de análise conceitual e tradução dizem respeito a essa outra “dimensão” da epistemologia. Essas tarefas implicam uma análise do tipo epistemológica. Para o profissional da informação é fundamental ter clareza quanto às próprias atitudes cognitivas frente a um documento a ser tratado.

Quando alguém formalmente enuncia do que um certo documento trata, essa enunciação é a expressão de uma crença do indivíduo a respeito do conteúdo discursivo/textual desse documento. Para que essa enunciação seja formulada é necessário antes identificar as “estratégias e métodos” usados pelo autor do documento em mãos para o estabelecimento e a afirmação das suas crenças.

Da mesma forma, essas estratégias e métodos a que Dancy (1995) se refere também servem para descrever aquelas que um indivíduo usa quando busca por informação “para conseguir novas crenças e filtrar as antigas”.

Saracevic (1996) exprime o problema fundamental da recuperação da informação nos seguintes termos:

“Considerando o problema da informação conforme definido, isto é, a explosão informacional, a recuperação da informação tornou-se uma solução bem sucedida encontrada pela ciência da informação e em processo de desenvolvimento até hoje. Como toda solução suscita seus próprios e específicos problemas, assim também a recuperação da informação e esses problemas estão contidos na concepção proposta por [Calvin] Mooers:

- a) como descrever intelectualmente a informação?
- b) como especificar intelectualmente a busca?
- c) que sistemas, técnicas ou máquinas devem ser empregados?

Embora tenham surgido outros problemas, mais específicos, esses três continuam fundamentais ainda hoje.” (SARACEVIC, 1996, p. 44).

Desses problemas fundamentais, dois possuem poderosas raízes epistemológicas. O objeto de estudo do profissional da informação é de natureza epistemológica. A ciência da informação é indissociável da filosofia da ciência da informação (ROBREDO (2007)).

A atividade documentária vai se desenvolver à medida em que for clarificado e formalizado o processo que denota *como* as crenças de um indivíduo em relação ao conteúdo de um documento são justificadas, e é a epistemologia que se ocupa com a justificação das crenças em geral.

Hjorland (2000), explorando o tema da prática, teoria e bases filosóficas da ciência da informação afirmou que:

“O verdadeiro desafio para a ciência da informação é, portanto, desenvolver conhecimento específico, que é *relativamente independente* do conhecimento do assunto, mas que não é uma abstração vazia. O Paralelo com a educação é desenvolver conhecimento sobre o ensino da matemática, música, química, etc. (a didática), que nem é idêntico ao conhecimento do assunto, nem é uma abstração vazia que não se relaciona aos problemas específicos de cada disciplina.” (HJORLAND, 2000, p. 506, grifo no original).

Essa descrição de Hjorland traduz bem a natureza do conhecimento epistemológico. Seu paralelo com a didática reforça ainda mais essa impressão. Estudamos as ciências durante os anos de nossa educação formal. Nesse período aprendemos seus principais conceitos, teoremas e leis. Isso não nos torna especialistas nessas diversas áreas, mas, nos permite a correta manipulação de seus constructos. O que afirmamos aqui com relação ao profissional da informação é uma extrapolação desse raciocínio.

## 2.4 O resumo na comunicação científica

A recuperação da informação e os produtos da atividade documentária evoluíram e se aperfeiçoaram. Contudo, enquanto as atividades específicas da indexação admitem um tratamento e uma análise de dados automática, a elaboração de resumos não admite. A produção de extratos pode ser feita adequadamente de uma forma automática (e isso nem é novidade veja, por exemplo, EARL (1970); BORKOWSKI e MARTIN (1975)), mas não a produção de textos.

Mueller (1997), por exemplo, não acredita que a situação daqueles envolvidos com o controle e a organização da atividade de comunicação científica tenha melhorado. Ao contrário, ela destaca que a pressão do *publish or perish* permanece de forma aguda, e não só isso, as tecnologias de informação e comunicação têm facilitado ainda mais a disseminação da informação de uma forma desordenada.

Um sistema de comunicação como o que implica a atividade científica, com uma carga de graves deveres sociais, precisa se ocupar com a correção dos desacertos que tornam o sistema menos eficiente. Bush (1945) nos lembra que:

“O conceito de Mendel das leis da genética ficou perdido por uma geração porque sua publicação não alcançou os poucos que poderiam compreendê-la e ampliá-la; e esse tipo de catástrofe, sem dúvida, está se repetindo entre nós, enquanto realizações verdadeiramente significantes ficam perdidas em meio a uma massa sem importância” (BUSH, 1945, p. 103).

O resumo, especialmente, possuindo as características de produto documentário que possui, vem sendo subproveitado em sua função na geração e disseminação de conhecimento científico por causa de sua qualidade contingente como produto documentário. Na “sociedade da informação” o efeito da atividade de condensação

informacional mais parece de *duplicação* de informação e, na verdade, o efeito dessa atividade deveria ser o de síntese e epítome do conhecimento divulgado.

O resumo cumpre uma função importante no acesso à informação, mas as imperfeições observadas em sua elaboração e no seu aproveitamento – na recuperação da informação – diminuem sua potencialidade.

Periódicos especializados em indexação e resumos surgiram a partir do século XVIII. Sua função era a identificação do conteúdo de publicações (de um campo definido por um recorte) e a produção de listagens de índices e resumos desses conteúdos. A TAB. 3 mostra alguns dos primeiros periódicos de indexação e resumos.

TABELA 3

Início dos serviços de indexação e resumos

1756	Journal Encyclopédique ou Universel (1756-93) Année Littéraire (1754-90)
1772	Esprit des Journaux Français et Étrangers (1772-1815; 1817-18)
1789	Annales de Chimie (1789-92; 1798-1815)
1797	Journal des Savants (1665-1792; 1797; 1816 até o presente)
1807	Neues Jahrbuch für Mineralogie, Geologie, und Paläontologie (1830-1949)
1816	Annales de Chimie et de Physique (1789-92; 1798-1815; 1816-1913)
1822	Law Journal Reports (1822-49)
1830	Pharmaceutisches Central-Blatt (1830-49)
1837	The Jurist: Weekly Periodical Containing Reports in All the Courts (1837-66)
1842	Nouvelles Annales de Mathématiques Journal des Candidats aux Écoles Polytechnique et Normale (resumos em 1842 e 1843)
1843	Medical News and Library (1837-42; 1843-79)
1843	Law Times Reports (London, 1843-1947)
1845	Die Fortschritte der Physik (1845-1918)
1853	The British Annual Abstract of Statistics (1840-53)

Fonte: Adaptado de BORKO e BERNIER, 1975, p. 30-31.

O aparecimento do resumo no periódico científico, acompanhando o artigo cujo conteúdo ele condensa, é um acontecimento que data do século XX. A TAB. 4 mostra alguns exemplos.

TABELA 4

Ano de início do uso de resumos junto dos artigos de periódicos científicos

Periódico	Ano de Fundação	Início do uso de resumos
Nature	1869	1970
Science	1883	1959
Physical Review	1893	1920
J. Biol. Chemistry	1905	1966
PNAS	1915	1969
Phys. Review Letters	1958	1967
J. Cell Biology	1958	1958
J. Molecular Biology	1959	1959
Biochemistry	1962	1962
Cell	1974	1974
Nucleic Acids Research	1974	1974
Molec./Cell. Biology	1981	1981

Fonte: HUCKIN, 1987 *apud* HARTLEY, 2000b, p. 274.

Define-se um resumo como “uma apresentação concisa dos pontos relevantes de um documento” (ABNT – NBR 6028:2003). As características de um bom resumo são “brevidade, exatidão e clareza” (LANCASTER, 2004, p. 113).

Cleveland e Cleveland (2001, p. 55-60) dividem os tipos de resumo em:

- a) indicativo – indica que informação significativa pode ser encontrada no documento. O resumo indicativo simplesmente descreve que tipo de documento está sendo resumido e do que ele trata; descreve as características temáticas gerais do documento sem comunicar informação específica sobre resultados ou conclusões.
- b) informativo – informa conteúdo do documento: apresenta tanto dado quantitativo e qualitativo quanto possível. O resumo informativo transmite as principais ideias e fatos do documento e contém dados importantes encontrados no próprio texto do documento resumido.
- c) crítico – apresenta conteúdo do documento e faz um julgamento de valor ou um comentário editorial sobre o documento.
- d) estruturado – é um resumo informativo que estrutura o conteúdo sob cabeçalhos como: Antecedentes, Objetivo, Método, Resultados, Conclusão.
- e) classificados por autor – resumos categorizados segundo sua autoria: seja o autor um especialista, seja um profissional da informação.
- f) classificados por uso – resumos categorizados de acordo com seu propósito: orientados à disciplina, ou orientados à missão.

Os resumos informativo/indicativo se destacam no processo de comunicação científica, sendo largamente utilizados por diferentes editores internacionalmente. Os resumos estruturados apareceram como alternativa na redação de resumos, apresentando grande potencial informativo.

O resumo estruturado foi adotado por grande número de periódicos na área médica, a partir do final da década de 1980 (AD HOC WORKING GROUP, 1987), (HARTLEY, 1994). Em 2005, o resumo estruturado foi incorporado como padrão da *Emerald*, que publica o *Journal of Documentation* na área de ciência da informação, conforme ilustrado na FIG. 6.

O resumo informativo/indicativo (FIG. 5) é normalmente redigido em um único parágrafo, com extensão variável entre 150 – 500 palavras, dependendo da política editorial adotada.

## Contextual Cocitation: Augmenting Cocitation Analysis and its Applications

**Alison Callahan and Stephen Hockema**  
Faculty of Information, University of Toronto, 140 St. George St., Toronto, Ontario M5S 3G6.  
E-mail: alison.callahan@gmail.com; steve.hockema@utoronto.ca

**Gunther Eysenbach**  
Centre for Global eHealth Innovation, Toronto General Hospital, R. Fraser Elliott Building,  
4th Floor, room #4S435, 190 Elizabeth Street, Toronto, Ontario M5G 2C4.  
E-mail: geysenba@uhnres.utoronto.ca

**In this work, a novel method of cocitation analysis, coined “contextual cocitation analysis,” is introduced and described in comparison to traditional methods of cocitation analysis. Equations for quantifying contextual cocitation strength are introduced and their implications explored using theoretical examples alongside the application of contextual cocitation to a series of BioMed Central publications and their cited resources. Based on this work, the implications of contextual cocitation for understanding the granularity of the relationships created between cited published research and methods for its analysis are discussed. Future applications and improvements of this work, including its extended application to the published research of multiple disciplines, are then presented with rationales for their inclusion.**

### Introduction

Traditional cocitation analysis is limited in the extent to which it can reveal the relationships between cocited resources in a document or a group of documents. The unit of analysis in cocitation analysis is the entire citing document, such that current methods of cocitation turn to the reference list of documents, rather than the rich information contained in the body of the citing document. In this work, it is proposed that one way to augment cocitation analysis, and subsequently its applications, is to consider the context in which a cocitation occurs as a measure of the strength of the cocitation. This proposal is explored by introducing a formulation of contextual cocitation strength and by applying this formulation to a set of BioMed Central publications to extract related publications. This work is framed by an exploration

of the current literature of cocitation analysis, citer motivation, and classes of citation. Future applications of contextual cocitation analysis include improving the literature selection process for research-motivated reading, studying the relationships between cited online resources, and, on a grander scale, rethinking the resolution of our existing methodological tools for exploring the relationships that exist among published works on a finer-grained basis.

### Related Work

#### Cocitation Analysis and Bibliographic Coupling

Author cocitation analysis and document cocitation analysis (ACA/DCA) are bibliometric analyses that seek to quantify the similarity between a pair of authors or documents (often relative to a larger set) by analyzing the number of cocitations in other documents to the pair. More specifically, document cocitation analysis quantifies the relationship between cocited documents (Small, 1973),<sup>1</sup> with the assumption that more frequently cocited documents exhibit a greater cocitation *strength* (Small, 1973, p. 265). Author cocitation analysis, as introduced by White and Griffith (1981), on the other hand, has as its unit of analysis the group of papers authored by a lone or primary author. The “author” in this case is considered to be the “body of writings by a person—and not the person himself” (White & Griffith, 1981, p. 163). Throughout the remainder of this work, the primary focus is on document cocitation analysis. (The implications of this work for author cocitation analysis, however, will be addressed.) Small (1974) also introduced a related bibliometric that he called “tri-citation,” which uses a third

Received June 12, 2009; revised September 17, 2009; accepted January 5, 2009

© 2010 ASIS&T • Published online 11 March 2010 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21313

JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 61(6):1130–1143, 2010

FIGURA 5 – Primeira página de um artigo publicado no JASIST com resumo informativo/indicativo

Fonte: JASIST, v. 61, n. 6.

O resumo estruturado (FIG. 6) tem uma aparência característica, dividido em parágrafos sob cabeçalhos. Sua extensão é bastante flexível sendo, em geral, um pouco maior do que o resumo informativo comum.

 The current issue and full text archive of this journal is available at  
[www.emeraldinsight.com/0022-0418.htm](http://www.emeraldinsight.com/0022-0418.htm)

## The *h*-index: a broad review of a new bibliometric indicator

Michael Norris and Charles Oppenheim  
*Department of Information Science,  
Loughborough University, Loughborough, UK*

The *h*-index

---

681

Received 19 June 2009  
Revised 5 January 2010  
Accepted 6 January 2010

**Abstract**  
**Purpose** – This review aims to show, broadly, how the *h*-index has become a subject of widespread debate, how it has spawned many variants and diverse applications since first introduced in 2005 and some of the issues in its use.  
**Design/methodology/approach** – The review drew on a range of material published in 1990 or so sources published since 2005. From these sources, a number of themes were identified and discussed ranging from the *h*-index's advantages to which citation database might be selected for its calculation.  
**Findings** – The analysis shows how the *h*-index has quickly established itself as a major subject of interest in the field of bibliometrics. Study of the index ranges from its mathematical underpinning to a range of variants perceived to address the indexes' shortcomings. The review illustrates how widely the index has been applied but also how care must be taken in its application.  
**Originality/value** – The use of bibliometric indicators to measure research performance continues, with the *h*-index as its latest addition. The use of the *h*-index, its variants and many applications to which it has been put are still at the exploratory stage. The review shows the breadth and diversity of this research and the need to verify the veracity of the *h*-index by more studies.

**Keywords** Information studies, Indexing, Information research  
**Paper type** Literature review

**1. Introduction**  
Scholars and departments frequently have the quality of their work assessed and ranked by their academic colleagues. The results of such assessments or rankings can lead to individual preferment, receipt of research funding or tokens of esteem.  
Apart from peer review, the impact of a scholar's work could, in part be assessed by the use of bibliometric indicators. Traditionally, for example, the major foci of citation studies have been the evaluation of academic departments, or of a countries' entire research output or of journals. A few citation studies have, however, focused on the performance of individuals by measuring their citation counts (Baird and Oppenheim, 1994; Brittain, 2000; Kalyane *et al.*, 2004). Generally, the assessment of an individual's impact by the use of citation analysis has been very limited, given the lack of a citation metric that might reflect a scholar's overall achievement.  
Recently, however, in 2005 Hirsch developed a new, simple and intuitively attractive measure of an individual's impact in his or her field – the *h*-index (Hirsch, 2005). Hirsch defined the index as follows:  
A scientist has index *h* if *h* of his/her  $N_p$  papers have at least *h* citations each and the other papers have no more than *h* citations each [...].

© The Authors, 2009.  
The authors wish to acknowledge the support of the AHRC in funding this research.



Journal of Documentation  
Vol. 66 No. 5 2010  
pp. 688-705  
Emerald Group Publishing Limited  
0022-0418  
DOI 10.1108/00220411011066790

FIGURA 6 – Primeira página de um artigo publicado no JD com resumo estruturado

Fonte: JD, v. 66, n. 5.

As normas técnicas desempenham um papel fundamental em atividades marcadas por grande subjetividade e introspecção como é o caso da elaboração de resumos. Só que quem se dedica a esta atividade no Brasil não dispõe de parâmetros adequados que orientem o seu trabalho. As normas propostas por entidades reguladoras nacionais definem aspectos do resumo relacionados à sua forma e deixam de definir aspectos relacionados à elaboração de seu conteúdo.

Tibbo (1992) observa que as normas técnicas internacionais relativas à redação de resumos são muito mais pertinentes às ciências exatas do que às humanidades. Ela discute a aplicabilidade das normas para redação de resumos e compara textos de psicologia, química e história. A autora acha evidências de que os resumos produzidos nos dois primeiros campos compartilham muitas características comuns e que cerca de 90% das sentenças nesses resumos entram nas categorias de conteúdo ISO 214. Para os resumos de história essa percentagem é inferior a 40%.

Tibbo conclui que tais normas e padrões são "bem concebidos para resumir a redação científica, mas, elas não coincidem com a realidade da literatura em história, ou, potencialmente, as características de outras redações não-científicas." (TIBBO, 1992, p. 47).

A norma internacional ISO 214 orienta que um resumo deva conter tanta informação quantitativa e qualitativa do documento original quanto possível. Essa orientação é feita assim, sem aprofundar ou sugerir métodos de como selecionar tal informação.

A ISO 214 faz uma distinção entre resumo e termos correlatos: comentário, extrato e epílogo e observa que, em muitas disciplinas, os leitores estão acostumados com resumos que informam o propósito, metodologia, resultados e conclusão relatados no documento original. No entanto, a norma também observa que a sequência de apresentação desses itens deve variar de acordo com a audiência pretendida.

Leitores interessados na aplicação de novos conhecimentos, segundo a ISO 214, se beneficiarão se os resultados e a conclusão aparecerem em primeiro lugar no resumo. Isso parece sugerir – como em Borko e Bernier (1975) – a ideia aqui defendida de que se uma informação específica pode ser encontrada imediatamente segundo uma

estrutura padronizada, isso deve ser feito. No caso da ISO 214, isso é feito padronizando os resumos de forma que resultados e conclusão figurem sempre no início dos resumos para determinada audiência, e no caso dos resumos estruturados isso é feito com informação específica que pode ser identificada imediatamente por meio de suas divisões e cabeçalhos (obrigatórios e optativos).

A ISO 214 não trata, propriamente, do resumo enquanto um texto. Sua proposta é a de explicar o que o resumo é e não tanto como redigir um resumo. A norma instrui o autor a iniciar o resumo com uma declaração central sobre o tema principal do documento (a menos que isso já esteja claro no título). Instrui também a usar parágrafo único para resumos curtos e vários parágrafos para resumos mais longos. A norma também determina o uso de verbos na voz ativa e na 3ª pessoa do singular. Tais orientações, ainda que sejam úteis, não orientam no sentido da seleção e concatenação de ideias, não promovem a formação de um texto.

A norma brasileira NBR 6028, por sua vez, tem como objetivo estabelecer os requisitos para a redação e apresentação de resumos embora seu título seja: “Informação e Documentação – Resumo – Apresentação”. Essa norma, de fato, fica melhor definida por seu título do que por seu objetivo. Ela consta de três itens: objetivo, definições e regras gerais de apresentação. Todo o seu conteúdo caberia numa única página de papel A4, escrita com fonte tamanho 10 e espaçamento entre linhas de 1,5.

A NBR 6028 define os termos: palavra-chave, resumo crítico, resumo indicativo e resumo informativo e, em seguida, trata da apresentação dos resumos. Ela regula que o resumo deve ressaltar objetivo, método, resultados e conclusão e explica que a ordem e extensão desses itens vão depender do tipo de resumo e do tratamento que cada item recebe no documento original. Aqui, de novo, essa orientação (apesar de se limitar a essas palavras sem maior aprofundamento) lembra a proposta de Kobashi (1997) defendida aqui.

A NBR 6028 orienta ainda o uso de um parágrafo único composto por uma sequência de frases concisas e não de enumeração de tópicos (sustentando de forma indireta a textualidade do resumo). Diz que o resumo deve iniciar com uma frase significativa

explicando o tema principal do documento com verbo na voz ativa e na 3ª pessoa do singular. E diz ainda que deve-se evitar símbolos pouco usados e fórmulas.

Conforme apontado na discussão sobre o ciclo informacional, a redação de resumos é uma função decisiva na iniciativa do controle bibliográfico e para a qualidade da comunicação científica.

Hartley e Benjamin (1998) afirmam que a importância dos resumos se acentuou com o crescimento da literatura científica:

“Na realidade, a natureza dos resumos alterou-se ao longo dos anos, na medida em que mais e mais artigos científicos passaram a competir entre si para atrair a atenção dos leitores. Hoje os leitores precisam compulsar e pesquisar mais do que o faziam no passado, e o resumo evolui continuamente como um portal de acesso à literatura científica.” (HARTLEY e BENJAMIN, 1998, p. 451).

A precária parametrização da atividade de elaboração de resumos reduz o potencial de sua contribuição para o controle bibliográfico. A geração de conhecimentos não depende da quantidade de informações a que se tem acesso, mas, de uma combinação específica do conteúdo de determinadas fontes de informação. A contribuição que o tratamento da informação e o controle bibliográfico podem prestar para o processo de comunicação científica vai além da filtragem de informações inúteis envolve, principalmente, a recuperação de informações úteis.

Os resumos tradicionais – informativos e indicativos – têm proporcionado pontos de acesso para seus originais e contribuído para a disseminação da informação, mas não sem efeitos colaterais indesejáveis. Esses resumos, frequentemente, omitem informações importantes; promovem redundâncias; se mostram enganosos e desencaminhadores no que diz respeito ao conteúdo informativo dos documentos que eles condensam.

Diversos autores em ciência da informação como Maizell, Smith e Singer (1971), Borko e Bernier (1975), Rowley (1988), Cleveland e Cleveland (2001), e Lancaster (2004) têm

prescritas orientações e regras quanto à redação de resumos na tentativa de corrigir os erros verificados nessa atividade ao longo dos anos em todas as áreas do conhecimento. Essas orientações versam sobre o formato e o conteúdo dos resumos e devem ser observadas durante a etapa de edição dos mesmos. Trata-se, no entanto, de orientações de caráter geral e cujo valor para a sistematização da atividade é pequeno.

Borko e Bernier (1975), em particular, apresentam uma lista de critérios para a avaliação de resumos:

- uma classificação global de qualidade, atribuída por avaliadores humanos;
- a medida em que as normas técnicas são respeitadas;
- a inclusão de informações importantes e a exclusão de informações sem importância;
- ausência de erros;
- coerência de estilo e legibilidade;
- previsibilidade da relevância;
- capacidade de servir como substituto do original;
- adequação como fonte de termos de indexação.

Os critérios de Borko e Bernier (1975) sinalizam claramente que a padronização da atividade de condensação ainda é muito incipiente e que essa atividade continua fortemente baseada experiência individual.

A deficiência de normas técnicas adequadas para a atividade de elaboração de resumos é muito bem exemplificada pelo estudo de Pitkin e Branagan (1998). Essas pesquisadoras trabalharam com autores que estavam submetendo artigos para publicação no *Obstetrics & Gynecology* e, nos resumos de alguns dos artigos submetidos, foram detectadas falhas. Orientações específicas foram dadas aos autores no sentido de corrigir as falhas detectadas. Isso foi feito no momento em que o artigo submetido era enviado de volta ao autor para alteração. Nenhuma correção foi

verificada. As pesquisadoras concluíram que instruções específicas dadas a autores para melhoramento de seus resumos são ineficientes.

Lancaster (2004, p. 105-106) argumenta que para certas finalidades um resumo estruturado é preferível a um resumo meramente informativo/indicativo. Lancaster usa o termo 'estruturado' de forma distinta daquela mais difundida e adotada aqui. Ele define um resumo estruturado a partir do gabarito hipotético mostrado abaixo:

TIPO DE IRRIGAÇÃO	TIPO DE SOLO	PRODUTOS	CONDIÇÕES CLIMÁTICAS	LUGAR	RESULTADOS
<hr/>					
<hr/>					
<hr/>					

**FIGURA 7 – Gabarito de um resumo estruturado segundo Lancaster**

Fonte: LANCASTER, 2004, p. 105.

No exemplo da FIG. 7, o resumidor é orientado a procurar especificamente os itens listados no gabarito. A elaboração do resumo consiste, então, em um exercício semelhante ao *estudo dirigido* muito usado no contexto de aprendizagem em sala de aula. Nesse exercício o estudante responde a uma lista de perguntas relativas a um mesmo texto didático. A estruturação de resumos, como sugerida pela FIG. 7, é claro, só seria aplicável ao processo de comunicação científica dentro de uma área temática específica o bastante para que os conceitos essenciais tratados nos diferentes documentos fossem os mesmos. O princípio de estruturação usado nesse tipo de redação de resumo, no entanto, é semelhante ao usado no resumo estruturado.

A tradição do uso dos resumos informativos/indicativos nos permite refletir e especular sobre alguns de seus aspectos. Lancaster (2004, p. 116) chama a atenção

para o fato de que em alguns serviços de indexação e resumos a primeira linha dos resumos é realçada como no exemplo:

---

A MIGRAÇÃO DE MÃO-DE-OBRA DE MOÇAMBIQUE PARA AS MINAS DA ÁFRICA DO SUL continua sendo um elemento importante nas relações econômicas entre estes países...

BROCKMANN, G. Migrant labour and foreign policy: the case of Mozambique. *Journal of peace research*. v. 22, 1985, p. 335-344.

---

Lancaster (2004) acha essa forma de apresentação mais atraente, e com potencial de captar a atenção do leitor sem grande esforço.

Da mesma forma, Borko e Bernier (1975, p. 69), apesar de defenderem que “não convém abrir parágrafos” em um resumo, e, de argumentarem que é desnecessário rotular cada parte do resumo – pois, “normalmente os leitores sabem qual é a parte que estão lendo” – esses mesmos autores propõem que “a colocação das conclusões em primeiro lugar satisfaz ao leitor e poderá dispensá-lo de continuar a leitura”.

Dessa última observação sobre a colocação das conclusões e da observação anterior de Lancaster (2004) extraímos um preceito diferente daquele fixado formalmente por Borko e Bernier (1975). A fácil localização de um item específico de informação é útil no contexto da busca de informação e deve ser empregada. A abertura de parágrafos com ‘rótulos’ possibilita isso. Ou seja, se em um determinado texto o leitor pode localizar imediatamente a informação desejada, isso deve ser feito.

O desenvolvimento conceitual alcançado pela atividade de elaboração de resumos ainda não conseguiu produzir um arcabouço teórico unificado. Muitos estudos sobre resumos estruturados, por exemplo, têm sido realizados:

Hartley (1998, 2003), Hartley e Benjamin (1998), Taddio *et al.* (1994), constataram em seus estudos que resumos estruturados contêm mais informação do que os resumos tradicionais. Hartley (2003), Hartley e Benjamin (1998) fizeram estudos envolvendo a

facilidade de leitura oferecida pelos dois tipos de resumo. Concluíram que os resumos estruturados são mais fáceis de ler. Hartley, Sydes e Blurton (1996) verificaram que os resumos estruturados oferecem uma perscrutação mais lógica e são mais fáceis de serem pesquisados. McIntosh (1997, 1999) investigou os efeitos que a estrutura textual dos resumos tem sobre a clareza da informação comunicada. Concluiu que os resumos estruturados facilitam a avaliação por pares no que diz respeito à apresentação de trabalhos em congressos. Haynes (1990, 1993) comprovou em suas pesquisas que os resumos estruturados têm, em geral, boa aceitação por parte de autores e leitores.

Tenopir e Jacsó (1993) analisaram três títulos de periódicos de índice e resumo, de interesse geral para leigos. O tipo de resumo usado nesses títulos é o tradicional. Os autores analisaram características dos resumos como: facilidade de leitura; extensão de parágrafos, sentenças e palavras; uso de voz passiva; e, uso de preposições. Um dado importante nesse estudo é que o melhor escore no quesito facilidade de leitura foi de 46 – em uma escala que vai de 0 até 100 – onde o intervalo de 70–100 indica um texto classificado de: razoavelmente fácil a muito fácil de ler. Os outros dois títulos aparecem com 39 e 37 pontos respectivamente. Na escala usada, o intervalo de 0–30 indica muito difícil de ler; o de 30–50, difícil de ler; o de 50–60, razoavelmente difícil; e o de 60–70, padrão.

Narine, Yee, Einarson, e Ilersich (1991) em seu estudo, encontraram que 52% dos resumos não descreviam explicitamente as variáveis do estudo. Ao descrever a seleção dos participantes, 79% falharam no uso de termos técnicos. Dos que informaram os resultados, 66% não forneceu dados para apoiar os resultados. Daqueles que informaram as conclusões, 86% não abordou as limitações do estudo.

Pitkin, Branagan e Burmeister (1999) levantaram indícios de que mais de 25% dos resumos de um periódico da área médica, mesmo após revisão feita por seus próprios autores, continham dados não encontrados no texto do artigo original.

Hartley (1999) verificou que os resumos estruturados, preparados por resequenciamento dos resumos tradicionais do periódico *Applied Ergonomics*, eram

mais longos, mais informativos e mais inteligíveis do que os originais. Ele encontrou resultados semelhantes em outro artigo (HARTLEY, 2003).

Hartley e Betts (2009) usaram um checklist para avaliar a qualidade de 100 resumos tradicionais publicados em 53 diferentes jornais de ciências sociais. A maioria desses resumos continha, efetivamente, informações sobre os objetivos, métodos e resultados dos estudos. No entanto, muitos não relataram detalhes sobre o tamanho da amostra, idade e sexo dos participantes, ou onde a pesquisa foi realizada.

Šauperl, Klasinc e Lužar (2008) partiram do fato de que o padrão ISO sugere que os resumos contenham os seguintes elementos: finalidade, métodos, resultados e conclusões. Diversos pesquisadores já desafiaram esta estrutura afirmando que diferentes disciplinas e culturas preferem informações diferentes. Essas afirmações foram apenas parcialmente apoiadas pelos resultados da pesquisa de Šauperl e colegas. Esses pesquisadores estudaram os resumos de trabalhos publicados em periódicos científicos eslovenos e internacionais.

Mai (2005) critica a indexação tradicional centrada no documento por não dar conta da complexidade envolvida na indexação. O autor propõe uma indexação centrada-no-domínio e defende que a força dessa abordagem é que ela, explicitamente, baseia a análise do documento na análise e entendimentos do domínio e dos usuários.

Pinto e Gálvez (1999) entendem que, no estudo sobre elaboração de resumos, dois aspectos merecem consideração especial: a *multi-dimensionalidade* – significando a complexidade estrutural, funcional e processual – e, a *interdependência*, ou, interação entre diferentes componentes, funções e processos. Para elas, a base científica da elaboração de resumos encontra-se em um pluralismo teórico e metodológico, sem o qual uma análise profunda deste processo complexo não é possível. “A investigação neste campo é caracterizada principalmente por dois aspectos aparentemente contraditórios: a unidade epistemológica e a diversidade. A fim de unir esses termos opostos, o contexto sistemático que os justifica deve ser levado em conta.” (PINTO e GÁLVEZ, 1999, p. 377).

Salager-Meyer (1991), em seu estudo sobre resumos de artigos médicos, chegou à conclusão de que metade deles possuía 'deficiência discursiva'. Essa denominação abrangia aspectos como dispersão conceitual (como resultados relatados em diferentes lugares do resumo) e omissão de elementos importantes (como a metodologia usada). O autor observa que a veiculação eficiente de informação científica é seriamente comprometida quando deficiências como essas interferem no processo.

Estudos ainda não foram feitos no que diz respeito aos efeitos mnemônicos dos resumos informativo/indicativo e dos resumos estruturados no leitor (HARTLEY, 2004).

Avanços na atividade de elaboração de resumos podem contribuir para a organização e disseminação da informação de uma maneira mais eficaz. Os resumos precisam, antes de mais nada, fazer uma representação *efetiva e singular* de seus originais, selecionando e sintetizando adequadamente conteúdo relevante.

Taddio *et al.* (1994), Haynes (1990, 1993), e Hartley (1996, 1998, 2000b) apontam para evidências de que o resumo estruturado, em relação ao informativo/indicativo, oferece uma leitura mais fácil e ágil; apresentam maior probabilidade de conter informações mais completas; têm um bom desempenho no que diz respeito à prevenção de omissão involuntária de conteúdo relevante; e, permitem uma obtenção de respostas, com menor margem de erro, a perguntas pontuais.

O levantamento dessas evidências oferece uma oportunidade para reflexão sobre o predominante uso de resumos informativos/indicativos nos periódicos nacionais em ciência da informação. Uma maior efetividade no processo de comunicação científica implica na adoção dos melhores instrumentos existentes.

A discussão sobre aspectos teóricos e aplicados da elaboração de resumos tem logrado um discreto espaço entre os pesquisadores da ciência da informação. Lancaster (2004, p.117) chama atenção para o fato de que "curiosamente, desde 1988, é provável encontrar na literatura médica um número maior de artigos sobre 'resumos' do que na literatura de ciência da informação". Tal observação nos permite questionar se a discussão sobre esse tema, em ciência da informação, já atingiu um patamar

epistemológico satisfatório, e, se existe um consenso esclarecido dentro da comunidade de pesquisadores quanto ao uso de resumos informativos/indicativos.

## 2.5 A elaboração de resumos

O processo de elaboração de resumos – apesar de sua prática muito difundida – ainda é pouco conhecido (KOBASHI, 1997), (MOLINA, 1995). Tal processo consiste em: 1) compreensão e seleção de informações textuais, e, 2) expressão do conteúdo selecionado na forma de um novo texto.

Há muito que se reconhece que a operação de seleção é decisiva para a elaboração de resumos, ao mesmo tempo, trata-se de uma operação desenvolvida ao sabor da experiência pessoal de cada profissional e pesquisador, cujo caráter tácito é um desafio para sua sistematização.

A seleção das informações para o resumo, entretanto, é realizada a partir de um objeto construído de forma rigorosamente estruturada: o texto. Esse arranjo de coisas sugere uma aproximação conceitual entre o modo como o texto é organizado e o modo como o próprio processo de seleção deve ser organizado e realizado.

Weisman (1972) defende que determinados elementos informacionais, dentro da estrutura do texto original, sejam salientados no resumo de acordo com o tipo de documento a ser condensado. A TAB. 5 abaixo ilustra seu argumento:

TABELA 5

Elementos a serem enfatizados no resumo de acordo com o tipo de documento

	Propósitos	Metodologia	Resultados	Conclusões
Relatórios de experimentos	✓	✓	✓	✓
Revisão de literatura	✓			✓
Exposições teóricas	✓			✓
Recensões críticas	✓			✓
Manuais		✓		
Livros-texto		✓		
Relatórios técnicos	✓	✓	✓	✓
Estudos de caso		✓	✓	✓
Anais de congresso		✓		

Fonte: Adaptado de WEISMAN, 1972, p. 231.

Nair Y. Kobashi, em seu artigo “Resumos documentários: uma proposta metodológica”, de 1997, propõe que a operação de seleção, no processo de elaboração de resumos, seja efetuada de acordo com as noções de texto, superestrutura e tipologia textual.

Segundo Kobashi (1997, p. 202), o texto é uma unidade de comunicação “organizada sintagmaticamente e dotada de coesão e coerência”. A classificação de textos já foi proposta de diferentes maneiras. Ela pode, por exemplo, obedecer a uma estrutura interna tipificando os textos em: descritivo, narrativo, dissertativo. Pode também caracterizá-los por sua finalidade: texto técnico, científico, didático, jornalístico, jurídico, político, etc.

Kobashi (1997) propõe uma análise de três variantes muito comuns de textos argumentativos: o texto canônico das ciências experimentais (a que preferimos designar como empíricos), o texto dissertativo e o texto expositivo.

## O texto empírico

Os textos empíricos são “elaborados com o intuito de expor metodicamente os resultados da observação de um fenômeno” (KOBASHI, 1997, p. 202). São textos que expressam o raciocínio de seu autor nas etapas que o conduziram “da observação de certos fatos empíricos ao enunciado de proposições denominadas de forma diversa: teses, hipóteses, interpretações, comentários, conclusões, explicações, etc.” (GARDIN, 1987, p. 4 *apud* KOBASHI, 1997, p. 202).

Kobashi (1997) define o texto científico como “uma unidade de comunicação do saber, dotada de certos elementos estruturais – ou superestruturais, na terminologia de Van Dijk e Kintsch (1983): o Tema, o Problema, a Hipótese, a Metodologia, os Resultados e a Conclusão” (KOBASHI, 1997, p. 202).

Ela explica que o tema é o assunto que se deseja desenvolver. Trata-se de uma ideia inicial abrangente a ser explorada e desenvolvida.

O problema, segundo Asti Vera (1980, p. 97), “é uma dificuldade, ainda sem solução, que é mister determinar com precisão, para intentar, em seguida, seu exame, avaliação crítica e solução”.

A hipótese é uma solução provisória – objeto de comprovação – para um problema considerado. É uma afirmação “de trabalho” ou operativa que deve se ajustar ao corpo de conhecimentos científicos existente. A hipótese pode ser explicativa/post-factum ou preditiva/ante-factum. O primeiro tipo organiza fatos e objetos dados como resultado do processo de investigação científica, buscando conciliá-los. O segundo tipo antecipa a forma como os fatos e objetos se organizarão e relacionarão ao final do processo investigativo.

A metodologia compreende “os procedimentos e operações que possibilitam a observação racional e controlada dos fatos, de modo a permitir a interpretação e a explicação adequada dos fenômenos observados” (KOBASHI, 1997, p. 203). Ela orienta tanto a coleta de dados quanto a sua análise.

Os resultados reúnem todos os fatos observados de uma forma ordenada e determinam se as hipóteses preditivas podem ser aceitas, e de que modo.

A conclusão é o desenredo do texto, onde são discutidos os resultados, sua aplicação e incorporação ao arcabouço teórico existente de uma forma coerente.

### **O texto dissertativo**

O texto dissertativo apresenta a seguinte superestrutura textual: tese, argumentos, e conclusão (CINTRA, 1987, p. 31).

Segundo Kobashi (1997, p. 203), a tese caracteriza-se pela apresentação de um ponto de vista. Os argumentos “qualificam positivamente o ponto de vista do autor e, simultaneamente, desqualificam um ponto de vista contrário”. A conclusão revalida o ponto de vista inicial apresentado.

### **O texto expositivo**

O texto expositivo apresenta o binômio problema/solução como categorias superestruturais básicas. O problema é entendido de forma semelhante à proposta por Asti Vera (1980, p. 97), e, a solução vincula causa e efeito de uma forma discriminável específica.

Em sua análise acerca desses três tipos textuais (empírico, dissertativo e expositivo), Kobashi (1997) apresenta um esquema para orientar a seleção de conteúdo durante a elaboração de resumos. Seu esquema é o seguinte:

TABELA 6

Esquema sugerido para a operação de seleção envolvida na elaboração de resumos

<b>Texto empírico</b>	Resumo Informativo	Resumo Indicativo
Tema	✓	✓
Problema	✓	✓
Hipóteses	✓	
Metodologia	✓	
Resultados	✓	
Conclusão	✓	✓

<b>Texto dissertativo</b>	Resumo Informativo	Resumo Indicativo
Tema	✓	✓
Tese	✓	
Argumentos	✓	
Conclusão	✓	✓

<b>Texto expositivo</b>	Resumo Informativo	Resumo Indicativo
Tema	✓	✓
Problema	✓	
Causas	✓	
Solução	✓	✓

Fonte: Adaptado de KOBASHI, 1997, p. 207.

Esse esquema reforça dois pontos: 1) a operação de seleção deve se concentrar e extrair informações de cada parte da estrutura discursiva, e, como consequência, 2) o resumo informativo deve exibir, como principal propriedade, um alto grau de similaridade e contiguidade com o texto original.

Nesse processo de seleção de conteúdo, a exclusão de certas partes do texto é feita mediante a atribuição de valor a outras partes que, combinadas, delimitam e definem aquilo que se pretende demonstrar com o texto como um todo.

Kobashi (1997) argumenta que há necessidade do estabelecimento formal de métodos, princípios e teorias para tratar a informação documentária. Ela critica as normas disponíveis que regulamentam essa atividade, afirmando que tais normas são ineficazes por duas razões: a) baseiam-se em regras intuitivas, na experiência ou no hábito, e, b) contemplam apenas a etapa de combinação de dados, desprezando o processo de seleção envolvido.

Para a autora, seleção e combinação são feitas com base em categorias específicas da superestrutura do texto científico (tema, problema, hipóteses, metodologia, resultados e conclusão). O uso da superestrutura – procedimento tão simples quanto fundamental – ajuda a garantir a qualidade e a completude das informações apresentadas em um resumo.

O resumo é um texto independente, mas, que mantém com o texto original uma relação de similaridade, suprimindo desse texto elementos como redundância. Em particular, o resumo de textos de caráter científico tende – partindo da extensa e detalhada linha argumentativa do texto de origem – para a expressão das premissas e da conclusão do texto condensado.

Na atividade de condensação informacional, a noção de texto é fundamental, pois a seleção de conteúdo para o resumo é feita com base na hierarquização de ideias e conceitos proposta e formalizada no conteúdo do texto original. A hierarquização de ideias e conceitos, aliás, transparece na própria forma do texto: através de títulos, subtítulos; marcações como sublinhado, negrito e itálico; e, na própria distribuição e ordenação de ideias nos parágrafos.

É importante observarmos que, na atividade documentária e na atividade de recuperação da informação, a representação só é possível se o texto original e o seu sucedâneo guardam uma estreita relação de similaridade e se existe congruência entre suas partes. Da parte dos descritores de um índice, a similaridade – ao nível da mensagem – é genérica, e, da parte do resumo, é específica. Além do mais, o texto do resumo guarda também uma relação de contiguidade com o texto original, ao nível da superestrutura.

Conforme Kobashi (1994), o resumo se caracteriza por representar *o documento*, e os descritores do índice se caracterizam por representar *o tema* do documento.

“A linguagem documentária [do índice], por ser um instrumento constituído por um léxico reduzido e sintaxe precária, não permite a produção de mensagens com a mesma virtualidade [sic] informacional dos resumos. Desse modo, para a indexação, interessa analisar o texto para determinar a sua macroestrutura de ordem mais geral, que será, em seguida, representada por palavras ou por expressões simples.” (KOBASHI, 1994, p. 108).

O resumo, por sua vez, envolve uma compreensão e uma seleção mais complexas, mantendo uma relação de semelhança e contiguidade com o texto original, ao nível da mensagem e da organização textual. O alcance expressivo do índice é menor, já que ele se relaciona com o original somente através de uma semelhança de superfície, ao nível não-específico da mensagem.

O resumo implica requisitos básicos de textualidade

“... por meio de operações comprometidas com a construção formal de uma mensagem dotada de coesão e coerência, enquanto o índice implica em elaborar representações constituídas por conjuntos de palavras justapostas ou relacionadas por uma sintaxe rudimentar...” (KOBASHI, 1994, p. 109).

Estudos sugerem que iniciativas como a de Kobashi (1997) podem ser implementadas com confiança. O estudo de Vinsonhaler (1966) é um deles. Esse pesquisador propôs que a ‘validade do conteúdo’ fosse avaliada com base em um julgamento do grau em que o resumo é ‘similar’ ao documento original, usando uma escala de similaridade.

Ele propôs um teste onde indivíduos examinam um documento e, em seguida, procuram identificar o resumo desse documento em uma folha impressa onde figuram também outros resumos. O objetivo desse teste é determinar em que medida o

resumo discrimina o documento que condensa, em especial quando os documentos, com os quais ele é confrontado, tratam do mesmo assunto ou de assuntos relacionados.

Vinsonhaler (1966) apresenta duas definições de 'validade comportamental': validade de conteúdo – o grau de semelhança de conteúdo entre um documento e seu resumo, e, validade preditiva – o grau de similaridade de resposta obtida de um documento e de seu resumo. A validade de conteúdo é medida através da ação dos participantes compararem diretamente documentos com resumos. A validade preditiva é medida através da correlação das respostas obtidas a partir dos documentos com as respostas obtidas a partir dos resumos. A validade preditiva foi avaliada positiva quando a validade de conteúdo era significativa.

Índices e resumos servem à representação de documentos. A representação proporcionada pelo índice é genérica, a representação que o resumo proporciona é específica. O uso do índice é mais adequado a uma abordagem exploratória. Através do uso do índice é possível encontrar documentos que têm boas chances de satisfazer uma necessidade informacional. O resumo, por outro lado, tem um caráter expositivo, discriminativo. Através do resumo é possível encontrar um documento que certamente vai satisfazer uma necessidade de informação.

É importante, em uma pesquisa que trata de resumos, reservar um momento para examinar mais de perto a relação entre índices e resumos, pois nos parece que um resumo do tipo indicativo não atende adequadamente os critérios e conceitualizações que regulam a atividade de recuperação da informação.

O resumo indicativo, como sucedâneo, não proporciona uma representação "interpretativa" de um documento (que possibilite uma interpretação baseada em dados efetivamente) pois não veicula informação específica.

Dentro do contexto de recuperação da informação, podemos argumentar que um conjunto de descritores extraídos de um índice e que descrevem bem um determinado documento, também pode ser adequadamente aplicado na descrição de vários outros documentos diferentes. No entanto, um resumo extraído de um banco de dados e que

descreve bem um determinado documento *não* poderá ser adequadamente usado para descrever nenhum outro documento.

O resumo *indicativo* guarda mais semelhança com o *índice* do que com o resumo e não conserva a utilidade nem de um e nem do outro. O resumo indicativo tem grandes chances de representar também outro documento além do original a partir do qual foi produzido.

Traçando um paralelo pouco sisudo, um relacionamento pela Internet não é fácil, mas tem mais chances de dar certo quando os interessados trocam fotos para se conhecer melhor. As chances desse relacionamento diminuiriam bastante se no lugar de fotos fossem trocadas apenas descrições genéricas como: “forte, moreno, simpático” e “alta, loira, olhos claros”.

É comum na comunicação científica o uso de resumos que veiculam conteúdo informativo e indicativo ao mesmo tempo. Essa escolha parece redundar em inconsistência no texto final assim produzido.

Nas palavras de Cleveland e Cleveland (2001):

“Índices e resumos não são fins em si mesmos, mas, instrumentos que auxiliam os usuários a encontrarem informações para resolução de seus problemas. Índices apontam para *possível* informação útil. Resumos permitem julgamento *eficiente* da relevância e pertinência daquela informação. [...] Resumos são também sucedâneos e complementam os índices resumindo o conteúdo efetivo do registro de conhecimento.” (CLEVELAND e CLEVELAND, 2001, p. 33, o grifo é nosso).

Um julgamento eficiente de relevância pressupõe similaridade e contiguidade textuais objetivas entre o resumo e o texto original (KOBASHI, 1994). O resumo indicativo não apresenta similaridade satisfatória com o original que representa. Na verdade, o resumo indicativo não atende a critérios específicos de textualidade. O tipo de referência que ele faz ao texto original o desautoriza como texto independente.

A tipologia aplicada aos resumos – apresentando como um dos tipos de resumo o indicativo – só não causa espanto por causa da precária parametrização e normalização disponível que orienta a redação de resumos no meio técnico-científico.

Na pesquisa aqui proposta, um resumo indicativo será considerado como deficiente, e, um resumo com parte de seu conteúdo expressa de forma indicativa, considerado com essa parte deficiente.

Poucos autores têm se envolvido com a pesquisa sobre resumos no Brasil. Além de Nair Y. Kobashi, Ana Cristina de O. Luz se dedica ao tema e discute a atividade de elaboração de resumos documentários a partir de propostas teórico-conceituais que, na verdade, influenciam as atividades documentárias como um todo.

Luz (1992) discute problemas encontrados em resumos informativos. A autora analisa resumos de trabalhos referentes à produção científica da EMBRAPA, mantidos em uma base de dados da própria entidade. Luz (1992) avalia os resumos informativos no que diz respeito ao cumprimento das normas EMBRAPA de qualidade para resumos, e identifica alguns desvios. A autora sugere maior atenção à regulação dos serviços de redação de resumos.

Achados consistentes com os de Kobashi (1997) e Luz (1992) foram observados por Fayol (1991), Guimarães (1993), Moreira González (1993), e, Molina (1995).

De modo geral, o descompasso entre as normas reguladoras e a prática de redação de resumos é uma constatação comum nas pesquisas sobre resumos. Outra constatação comum é a que se refere à resistência que a atividade mostra à sistematização.

No que tange à complexidade e interdisciplinaridade envolvidas no processo de elaboração de resumos, constructos teóricos de outras disciplinas – como o oferecido pela psicologia cognitiva, por exemplo – são apontados como essenciais para o entendimento da atividade de resumir. Apesar disso, as mudanças propostas nesses casos têm privilegiado a área de análise documentária, reforçando a adoção de metodologias, práticas e instrumentos próprios da área, na elaboração dos produtos informacionais (MOLINA, 1995); (LUZ, 1992).

Orientações gerais para elaboração de resumos têm sido propostas. Diversos autores avançaram esquemas e modelos na tentativa de sistematizar essa atividade. Muitas dessas propostas se baseiam unicamente na experiência pessoal do autor ou em suas reflexões sobre o trabalho de outros teóricos.

O modelo de Cremmins (1996), por exemplo, pode ser descrito em três etapas:

Leitura Exploratória-a-Recuperatória.

Leitura Proativa-a-Inventiva.

Leitura Conectiva (Valor-ao-Significado).

A Leitura Exploratória-a-Recuperatória é feita “de fora para dentro”, pois o leitor ainda não está familiarizado com o conteúdo. O leitor escrutina exploratoriamente o texto identificando passagens de potencial interesse para inclusão no resumo. Enquanto escrutina, o leitor faz anotações às margens do texto nas partes que contêm informação sobre propósito, métodos, conclusões e recomendações.

Na leitura Proativa-a-Inventiva o leitor indexa mentalmente os temas primário e secundário enquanto relê as partes que contêm informação sobre propósito, métodos, conclusões e recomendações. Qualquer outra informação que o leitor considere temática deve ser incluída no resumo. Em textos de natureza empírica, o leitor deve reler os resultados de forma especial. O leitor deve extrair desse tipo de texto as conclusões e recomendações mais relevantes.

Na leitura Conectiva (Valor-ao-Significado) o leitor analisa seu esboço do resumo, procurando por erros no estilo e no conteúdo. Ele procura agregar ao resumo valores como abrangência, precisão, unidade, continuidade, seletividade, relevância, universalidade e conveniência.

Laurent (1985 *apud* GIASSON e FRIAS, 1993, p. 115) afirma que a redação de resumos consiste na “reescrita de um texto com um triplo objetivo: a conservação da equivalência informativa, a concretização de uma economia de meios de significação e a adaptação a uma nova situação de comunicação.”

Molina (1995) divide o processo de sumarização em quatro passos-chave: leitura-compreensão; seleção; interpretação; síntese/descrição analítica.

Serafini (1987, p. 188) preceitua quatro regras para a redação de resumos: cancelamento; generalização; seleção; e, construção.

Marcuschi (2001), em sua análise da textualidade oral e escrita e retextualização, propõe um conjunto de operações cujo objetivo é elucidar as transformações que comumente ocorrem na passagem de um texto oral para um texto escrito. O próprio autor sustenta que o processo de retextualização ocorre de forma semelhante em outros contextos, como na passagem de um texto escrito para outro.

Para Marcuschi (2001), ler um texto implica em resumir/sumarizar, tendo como efeito a produção de um novo texto (escrito ou não). Esse processo é designado retextualização. A construção de uma 'macroestrutura' orienta todo esse processo.

O leitor, durante seu contato com um texto, intuitivamente, constrói uma macroestrutura dele de forma a hierarquizar as ideias. Para isso, o leitor sumariza o texto como forma de reter as ideias principais e suas relações recíprocas. Isso envolve estratégias de apagamento de seções textuais (com seleção de proposições relevantes), e, estratégias de substituição de proposições (com generalização temática).

Marcuschi (2001) explica que:

“A redução de informações de um texto é operada a partir de “macrorregras”, que são princípios organizacionais de processamento de informações complexas com os quais se formula a macroestrutura de um texto. Ao construir a macroestrutura de um texto, o leitor elimina detalhes, seleciona os tópicos mais relevantes, reformula-os sinteticamente em proposições básicas e constrói seqüências mais globais através de transformações que se realizam via operações de substituição, acréscimo, reordenação ou condensação de elementos dos fragmentos do texto.” (MARCUSCHI, 2001, p. 74).

O processo de elaboração de resumos, então, pode ser considerado como um processo de 1) leitura do texto objeto de resumo, 2) seleção dos conceitos que tematizam o texto, e, 3) redação do resumo propriamente dito.

1) O primeiro passo consiste em uma leitura ativa para identificação de informações (numa atividade de pré-seleção), e passiva para a compreensão (CREMMINS, 1996, p. 28).

2) O segundo passo é marcado por um duplo processo: supressão e seleção.

- **supressão** de ideias e trechos redundantes ou acessórios ao tema.

- **seleção** das ideias e trechos centrais para a lógica discursiva adotada no texto original, evidenciada em sua superestrutura.

3) No último passo, há uma generalização da linha argumentativa usada no texto original e uma combinação das ideias ali presentes. Trata-se de um processo de “re-tematização” que é expressa por meio de um texto coeso e coerente.

A atividade intelectual envolvida no processo de sumarização é a mesma empregada tanto na compreensão quanto na produção de textos (MOLINA, 1995). Trata-se de uma atividade de análise e síntese básica e essencial para a vida dentro de uma sociedade que cria e recria representações de suas experiências e de seu conhecimento usando a palavra escrita.

A sumarização, além disso, está presente em diversos campos da ação humana, como educação formal, jornalismo, comunicação científica, mercado editorial, documentação, entre outros. Apesar disso, todo o conhecimento e toda a história de técnica e prática que temos sobre sumarização ainda não foram consolidados teoricamente de maneira satisfatória.

Dada a complexidade de fatores que cerca o processo de elaboração de resumos, autores como Hutchins (1987) e Molina (1995) chamam a atenção para a necessidade de mais estudos sobre esse processo. Tais estudos devem ser integrados em um arcabouço teórico abrangente, interdisciplinar, pois a elaboração de resumos é objeto

de estudo e interesse de diversas disciplinas. Segundo esses autores, a própria complexidade envolvida serve como um registro da sua interdisciplinaridade.

Parece interessante fazer aqui uma observação acerca do caráter complexo do processo de sumarização. A leitura é o primeiro passo do processo de sumarização. Cremmins (1996, p. 28) é quem propõe a regra: “ler ativamente para identificar informação para o resumo, e passivamente para compreensão” para o processo de sumarização. Sua afirmação é paradoxal e provocativa.

Durante a leitura de um texto, o leitor identifica uma ideia de interesse na trama simbólica constituída pelo texto e procura por uma outra com a qual ela se relacione diretamente e, então, procura entender distintamente a natureza dessa relação entre elas. Esse exercício prossegue numa atividade de síntese cujo desencadeamento resulta numa “ideia geral” do texto.

Essa noção é reforçada por experiências como a da lógica simbólica. Alfred N. Whitehead, matemático e lógico que fez grandes contribuições ao estudo da lógica simbólica, afirmou que:

“...com a ajuda do simbolismo podemos efetuar quase mecanicamente, por meio da vista, transições no raciocínio as quais exigiriam, sem aquele, o uso das faculdades superiores do cérebro.”  
(WHITEHEAD e RUSSELL, 1950 *apud* COPI, 1978, p. 226).

Irving M. Copi, em seu livro “Introdução à lógica” recorda essa mesma passagem e acrescenta:

“Deste ponto de vista, chegamos à conclusão bastante paradoxal de que não diz respeito à lógica o desenvolvimento de nossas faculdade de pensamento, mas o desenvolvimento de técnicas que nos habilitem a avançar sem ter que pensar!” (COPI, 1978, p. 226).

Apesar do tom bem-humorado de Copi (1978) sua intenção é argumentar em favor da lógica simbólica, promovendo suas virtudes.

O uso de símbolos para a representação de ideias mais complexas (como o uso da construção: -----  $p \vee q, \sim p \rightarrow q$  ----- para a representação de um argumento na forma de um 'silogismo disjuntivo') permite que as relações propostas sejam melhor e mais facilmente apreciadas. Em consequência disso, as próprias ideias são melhor apreendidas. Essa melhor apreensão ocorre porque compreender as relações e efeitos de uma coisa qualquer é conhecer mais sobre essa coisa.

As citações de Whitehead (1950) e Copi (1978) – à luz da exposição feita aqui – deixam ver que existe uma conformidade entre a percepção dos teóricos em lógica e aquela regra proposta pelo documentalista (CREMMINS, 1996).

### 3. Metodologia

Na pesquisa aqui proposta a abordagem se concentrou, especificamente, na análise da produção científica de pesquisadores em ciência da informação.

Comparou-se a produção veiculada no JD com a produção veiculada no JASIST, no ano de 2010. Isso foi feito, primeiramente, sob a perspectiva de identificar-se o tipo textual e a superestrutura dos artigos publicados e sua influência na produção dos resumos publicados.

O corpus da pesquisa constitui-se de 84 resumos: 42 tradicionais (informativos/indicativos) – publicados junto aos artigos originais no JASIST – e, 42 estruturados – publicados junto aos artigos originais no JD.

Em um primeiro momento, fez-se uma análise pormenorizada dos próprios artigos publicados para definição da estrutura temática e da superestrutura textual adotada em cada um, conforme aplicável ou não o esquema de Kobashi (1997).

Em um segundo momento, verificou-se se o resumo contemplava conteúdo de todos os itens da superestrutura do texto original que resumia. A superestrutura de cada artigo do corpus – por instrumentalizar a contiguidade textual com o resumo – determinava conteúdo relevante, juntamente com a similaridade textual.

Finalmente, procedeu-se a uma comparação entre resumos informativos/indicativos e resumos estruturados para a verificação do desempenho de cada um no que diz respeito à capacidade de representação de informação relevante, dentro do processo de comunicação científica.

Considerou-se como unidade de análise o próprio esquema proposto por Kobashi (1997). Durante o exame de cada artigo científico do corpus da pesquisa era verificado se o artigo correspondia a um dos três tipos textuais abrangidos pelo esquema: empírico, dissertativo e expositivo. Feito isso, procedia-se ao exame do respectivo

resumo no sentido de verificar se foi feita representação de conteúdo relevante de cada item da superestrutura textual usada.

Segundo a proposta de Kobashi (1997), uma representação efetiva – no caso específico de textos escritos – só é possível se original e sucedâneo guardam uma relação de similaridade e contiguidade. A identificação do conteúdo do resumo foi feita com base na hierarquização das ideias observada no texto original. A hierarquização de ideias e conceitos transparece na própria forma do texto, conforme destacamos em nossa discussão sobre elaboração de resumos.

Parte-se do pressuposto de que os resumos estruturados proporcionam uma representação distinta e de maior quantidade de informação relevante. Por causa disso, eles proporcionam uma representação mais eficaz do texto a partir do qual foram produzidos.

Optou-se por estudar a comunicação científica formal. Os periódicos escolhidos têm reconhecimento internacional, sendo listados na base de dados Qualis da Coordenação para Aperfeiçoamento de Pessoal de Ensino Superior (CAPES). O JD possui um Índice H de 34 e um Fator de Impacto de 1.67 (2009), e, o JASIST possui um Índice H de 58 e um Fator de Impacto de 2.3 (2009).

O Índice H e o Fator de Impacto servem como referências para medir o impacto de determinada produção científica, são normalmente aplicados à produção bibliográfica de um pesquisador individual, mas também podem ser estendidos para se obter uma percepção do impacto dos próprios periódicos que veiculam as mais diversas produções.

O Índice H só leva em consideração os artigos citados de determinado periódico, e não todo o conjunto publicado. No caso do JD, esse periódico publicou 34 artigos que receberam, pelo menos, 34 citações cada um. O JASIST publicou 58 artigos que foram citados, pelo menos, 58 vezes cada um. Esses índices são considerados expressivos dentro do contexto competitivo e numericamente vasto da comunicação científica.

O Fator de Impacto identifica a frequência com que o artigo médio de um periódico é citado em um determinado ano. Esse número pode ser usado para avaliar ou comparar

a importância relativa de um periódico com outros do mesmo campo ou ver com que frequência os artigos são citados para determinar quais periódicos são melhores para os interesses de determinado pesquisador. O artigo médio do JD, só no ano de 2009, recebeu 1,67 citações, e, o artigo médio do JASIST recebeu 2,3 citações.

É claro que, sobre esse assunto de relevância e impacto na comunidade científica, alguns aspectos precisam ser levados em consideração. Dentre eles, deve-se observar que o JD não publica artigos com forte abordagem tecnológica como o JASIST faz. Tais artigos vão ser de muito interesse para autores em ciência da computação, por exemplo. Essa diferença afeta significativamente a interpretação que um estudioso pode fazer daqueles números, seja com relação ao interesse que ele tenha em uma comunidade específica (de colegas pesquisadores), seja com relação a uma abordagem que valorize a perspectiva interdisciplinar, por exemplo.

### ***Journal of Documentation***

O JD tem uma longa história na área de biblioteconomia e ciência da informação e constitui um fórum para a divulgação de artigos acadêmicos, relatórios de pesquisa e resenhas críticas em ciência da informação. Ele fornece uma ligação entre investigação empírica, aspectos teóricos e prática profissional e busca informar e promover o avanço de todas essas frentes de atuação.

Todos os artigos submetidos passam por uma dupla revisão cega por pares (*double blind review*). O JD tem a perspectiva única de se concentrar exclusivamente em teorias, conceitos, modelos, estruturas e filosofias da ciência da informação. O jornal também publica relatórios de pesquisas diversas, desde que estas tenham significado amplo e de relevância para a ciência da informação. O JD também publica artigos sobre metodologia de pesquisa, história da informação, disciplinas de informação – incluindo questões educacionais, currículo e conexões com outras disciplinas – e que façam uma relação entre o estudo acadêmico e a prática profissional.

Algumas áreas de abrangência são:

- \* Ciência da informação, biblioteconomia e disciplinas relacionadas;
- \* Informação e gestão do conhecimento;
- \* Informação e organização do conhecimento;
- \* Busca e recuperação da informação;
- \* Comportamento informacional;
- \* Informação e letramento digital.

### ***Journal of the American Society for Information Science and Technology***

O JASIST tem sido publicado continuamente desde 1950. Editado por Blaise Cronin, o JASIST publica relatórios de pesquisa e desenvolvimento em uma ampla gama de temas e aplicações em ciência da informação e tecnologia. O JASIST começou como American Documentation. Em 1970, tornou-se JASIS. O nome mudou novamente em 2000 para o nome atual, JASIST.

O JASIST é, provavelmente, o principal fórum internacional para a pesquisa avaliada por pares em ciência da informação. Ele tem proporcionado a liderança intelectual através da publicação de pesquisas originais que focam a produção, a descoberta, a gravação, armazenamento, representação, recuperação, apresentação, manipulação, difusão, utilização e avaliação de informações. Suas pesquisas abordam também as ferramentas e técnicas associadas a esses processos.

O JASIST apresenta trabalhos de natureza empírica, experimental, etnográfica, conceitual, histórica, sócio-técnica, político-analítica, e, teórico-crítica. O JASIST também apresenta artigos de revisão em profundidade denominados: "Avanços em Ciência da Informação". Os "Avanços em Ciência da Informação" são uma série de artigos que continuam a tradição estabelecida pelo *Annual Review of Information Science and Technology* (ARIST).

## 3.1 Corpus da pesquisa

O JD é um periódico bimestral. Cada um de seus números bimestrais apresenta 7 artigos científicos. O JD adota resumos estruturados em suas publicações desde 2005. A norma que regula a elaboração desse tipo de resumo está claramente expressa em sua política editorial e nas orientações aos autores. No JD foram analisados todos os 42 artigos e resumos publicados nos seis números do volume 66, em 2010.

O JASIST tem uma periodicidade mensal. Cada número apresenta entre 10 e 15 artigos científicos. O JASIST adota resumos informativos/indicativos em suas publicações. Foram analisados 42 artigos e resumos selecionados dos seis primeiros números do volume 61, publicados em 2010.

O corpus da pesquisa foi determinado com base em um recorte de tempo. Entendeu-se que a análise de um exercício administrativo – de janeiro a dezembro de 2010 – dos periódicos selecionados possibilitaria um horizonte atual da prática documentária no que diz respeito aos resumos.

O ano de 2010, também, permite observar os efeitos e repercussões de cinco anos de uso de resumos do tipo estruturado em ciência da informação, pois o JD foi pioneiro na adoção desse tipo de resumo na área. Além disso, como o JD e o JASIST se situam em um mesmo quadro cultural anglo-saxão (o JD de origem britânica e o JASIST de origem americana) é de se esperar que eles se influenciem mutuamente de forma mais próxima.

O recorte temporal definido também abarca eventos importantes para a área que dialogam com a produção acadêmico-científica daquele ano. O encontro da ASIST (22 a 27 de outubro de 2010 – Pittsburgh, EUA) que é responsável pela publicação do JASIST, e, a EURO INFORMATION ARCHITECTURE SUMMIT (setembro de 2010 – Paris, França).

No encontro da ASIST, um trabalho sobre resumo foi apresentado, seu título era: “Human Abstracts, Machine Summaries, Cyborg Solutions?”. Nos encontros mencionados, diversos trabalhos em indexação e classificação foram apresentados. É

bom destacar que tais eventos indicam tendências e preocupações correntes entre os pesquisadores.

Considerando um nível de significância de 0.05, o corpus de 84 resumos determinado pelo corte temporal tem um poder estatístico de 0.7851151 para a detecção de uma correlação moderada, usando um teste qui-quadrado com grau de liberdade igual a um.

Salienta-se que o recorte temporal abrange uma iniciativa autônoma e uma aplicação real do uso de resumos estruturados em publicações de caráter científico. Em alguns estudos sobre resumos – como Hartley (1999, 2003) – resumos informativos foram vertidos para o formato estruturado pelo próprio autor do estudo para fins de análise.

## **3.2 Procedimentos de coleta dos dados**

A coleta de dados para a avaliação dos artigos e resumos seguiu um procedimento de compilação simples reunindo integralmente os artigos e resumos publicados no volume 66 do JD, e, reunindo artigos e resumos selecionados do volume 61 do JASIST.

Os artigos do JASIST foram selecionados de forma a se obter equivalência na operação de comparação, já que o número anual de artigos publicados pelo JASIST é quatro vezes o número anual de artigos publicados pelo JD. Da mesma maneira, como o JASIST – em acordo com os objetivos expressamente enunciados em sua política organizacional – publica grande número de artigos especificamente das áreas de ciência da computação e tecnologia, procuramos excluir de nosso corpus esses artigos para uma melhor comparação com os conteúdos publicados pelo JD.

A operação de coleta de dados foi realizada entre dezembro de 2010 e janeiro de 2011, pois o Portal de Periódicos da CAPES não permite o download consecutivo de todos os artigos científicos de um mesmo número (de qualquer periódico que seja) num curto espaço de tempo.

### 3.3 Procedimentos de análise dos dados

Os artigos foram examinados com o objetivo de se fazer seu levantamento temático e o levantamento das superestruturas textuais neles usadas (KOBASHI, 1997). Buscou-se no processo de análise desses artigos e de seus resumos ampliar o entendimento da atividade de retextualização (MARCUSCHI, 2001).

Observou-se que o emprego de linguagem formal e objetiva nos resumos analisados não se traduziu necessariamente em clareza e transparência na comunicação. Entretanto, de forma geral, a ordem dos fatos como eles aparecem no resumo informativo não se afastou muito da ordem em que eles eram apresentados no texto original (dado que o resumo estruturado se caracteriza por uma ordem pré-estabelecida para a apresentação dos fatos).

Todavia, salientamos que, apesar da ordem de apresentação dos fatos e conceitos nos resumos informativos ter apresentado alguma regularidade, isso não implica em eficiência no processo de comunicação científica.

O resumo é um item documentário e sua relação com o usuário é distinta da relação entre esse usuário e o artigo original que o resumo condensa. O usuário, no processo de busca por informação, consulta um número maior de resumos do que de artigos. Esse processo iterativo é mais eficiente se for padronizado, pois ele envolve características mecânicas.

O usuário se encontra em um processo de busca de informação: trata-se de um processo de inspeção ativa para identificação de informação (CREMMINS, 1996). A inspeção de itens estruturalmente semelhantes se torna mais ágil, e esse não é o caso com resumos informativos tradicionais.

Segundo pressuposto na análise aqui proposta, palavras/frases do resumo deveriam ser substanciadas por conteúdo relevante/detalhado no texto original. Isso ocorreu, como veremos adiante, com os resumos (dos dois tipos) que atenderam aos critérios de textualidade e à superestrutura textual.

As investigações de Kobashi (1997) e Marcuschi (2001) chamaram nossa atenção para o fato de que durante uma leitura o indivíduo concentra-se na atividade de significação e, internamente, ele experimenta tematizar e relacionar os diversos conteúdos de cada item da superestrutura usada no texto. Dessa forma, ao final do processo de leitura é possível enumerar os principais conceitos e delimitar com segurança do que o texto trata.

No processo de leitura o texto é dividido em fragmentos abrangidos por um mesmo tópico e cada conjunto desses fragmentos irá constituir uma unidade de nível mais alto. Várias dessas unidades, por sua vez, formarão uma unidade semântica de nível ainda superior e assim por diante até que se forme na mente do leitor a “noção de texto” (KOCH, 1992).

Na análise dos artigos científicos do JD e do JASIST foi identificado o tipo do texto em mãos: se empírico, dissertativo, expositivo ou outro. A partir daí, foram identificados os itens da superestrutura adotada pelo autor, conforme Kobashi (1997):

- no texto empírico: tema, problema, hipóteses, metodologia, resultados e conclusão.
- no texto dissertativo: tema, tese, argumentos e conclusão.
- no texto expositivo: tema, problema, causas e solução.

Terminada essa etapa, os apontamentos feitos durante a análise do artigo científico foram usados para a análise do seu resumo. Procurou-se verificar se informação objetiva (não meramente indicativa) correspondente a cada item da superestrutura do texto original foi condensada. Finalmente, foi feita a comparação do desempenho de dos resumos informativos e dos resumos estruturados quanto à eficácia de representação.

## 4. Resultados

Dos 42 artigos analisados no JASIST, 38 relatavam estudos *empíricos* e usavam a superestrutura textual do tipo correspondente; 2 artigos eram do tipo *dissertativo*; e outros 2, do tipo *expositivo*. Os resumos correspondentes eram informativos e informativo-indicativos.

Dos 42 artigos analisados no JD, 32 se caracterizavam como artigos *empíricos*; 8 eram do tipo *dissertativo*; e outros 2, do tipo *expositivo*. Os resumos correspondentes eram estruturados.

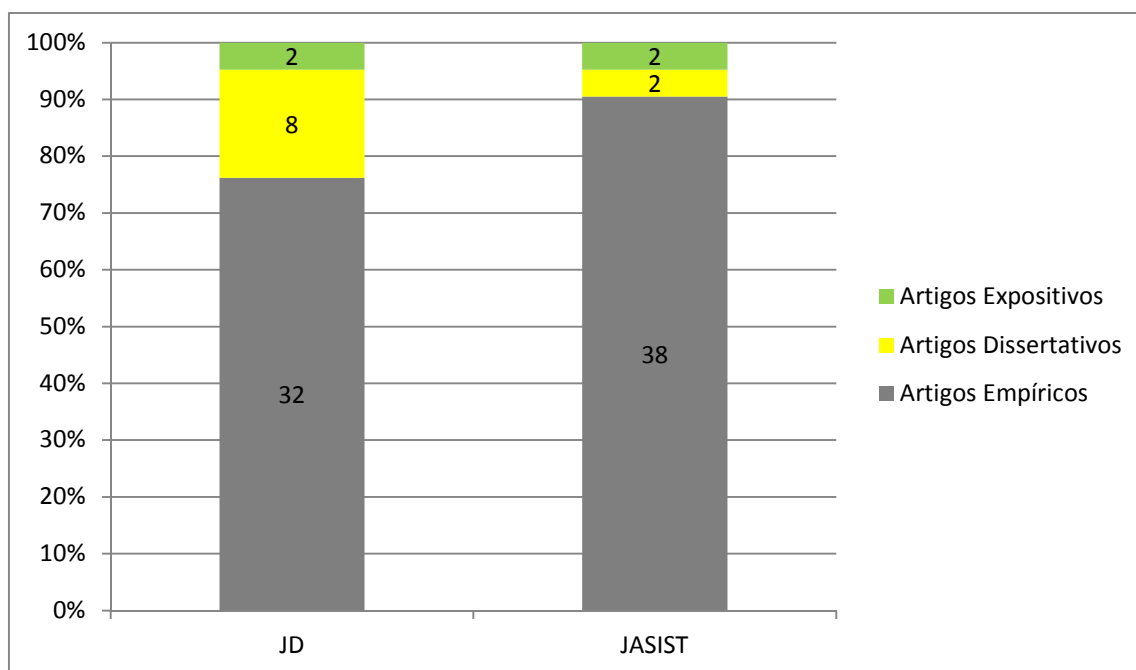


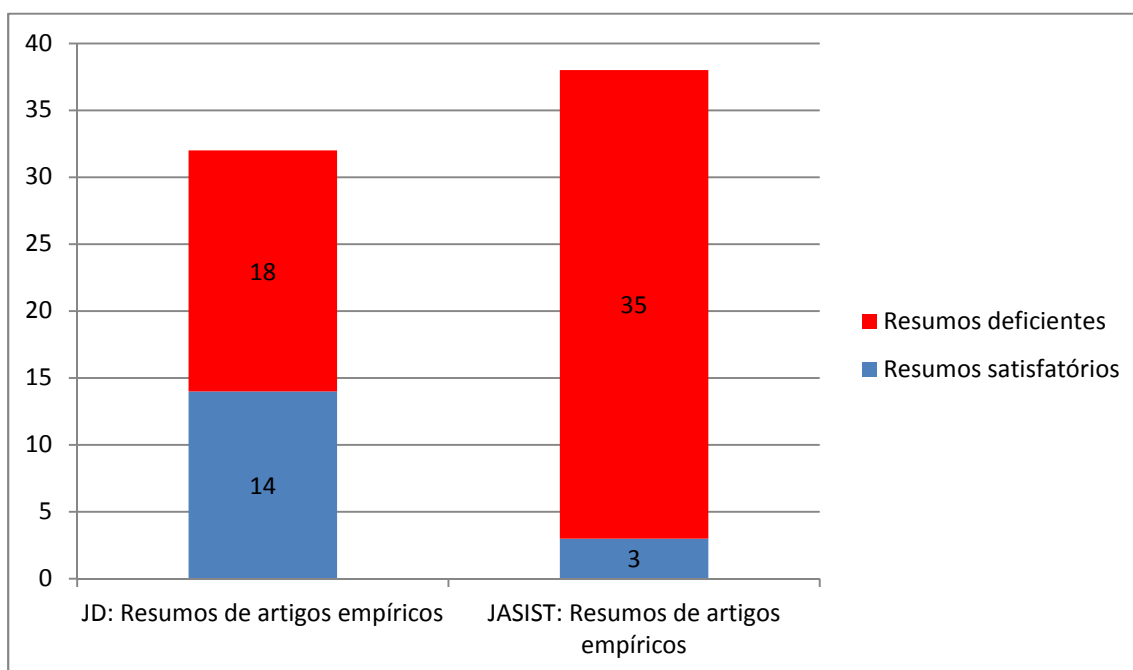
GRÁFICO R1 – JD e JASIST quanto aos tipos de artigos publicados em 2010

Nesses periódicos de tradição consolidada, o grande número de trabalhos empíricos – relativamente aos trabalhos de abordagem dissertativa e expositiva – reflete o caráter aplicado de uma área que se ocupa com um recurso amplamente difundido e empregado: a informação. Em ambos os periódicos essa constatação foi verdadeira.

A proposta de Kobashi (1997) demonstrou ser facilmente aplicável tanto na análise dos artigos quanto na avaliação dos resumos, facilitando efetivamente o processo de decomposição do texto em fragmentos coerentes e ordenados, fiéis à estrutura temática.

As superestruturas textuais são comumente utilizadas no discurso científico, e são usadas com competência por pesquisadores e acadêmicos (que possuem anos de trato com o texto escrito). A apropriação que um determinado autor faz de uma superestrutura para a construção do seu texto ocorre de forma quase natural, dadas as características convencionais e bem conhecidas inerentes à superestrutura.

A análise feita com o recurso das superestruturas textuais revelou inconsistências e deficiências nos resumos dos dois tipos. Essas falhas produziram efeitos na veiculação de informação relevante e na representatividade ensejada pelos resumos enquanto produtos documentários. Por outro lado, observou-se que todos os casos de resumos satisfatórios abrangiam resumos de artigos do tipo empírico:



**GRÁFICO R2 – Ocorrências de resumos satisfatórios no JD e no JASIST**

Quanto à veiculação “completa” de informação relevante – efeito da similaridade e contiguidade textuais – foi observado que 17 dos 84 resumos analisados se

apresentaram satisfatórios. O índice geral não é satisfatório (20,24%), mas, comparativamente, os resumos estruturados se saíram melhor que os informativos:

#### DADOS CONSIDERANDO SOMENTE RESUMOS DO TIPO EMPÍRICO

3/38 ou 7,89% dos resumos informativos proporcionaram uma representação satisfatória em relação ao original.

14/32 ou 43,75% dos resumos estruturados proporcionaram uma representação satisfatória em relação ao original.

#### DADOS CONSIDERANDO RESUMOS EM GERAL

3/42 ou 7,14% dos resumos informativos proporcionaram uma representação satisfatória em relação ao original.

14/42 ou 33,3% dos resumos estruturados proporcionaram uma representação satisfatória em relação ao original.

Comparando esses resultados com os obtidos por Hahs-Vaughn e Onwuegbuzie (2010), que trabalharam com um corpus de tamanho semelhante ao definido aqui, vemos que o número de resumos deficientes que eles encontraram (44,3 %) é bem menor do que o encontrado aqui (79,76 %). É verdade, no entanto, que aqueles autores não consideraram, aprioristicamente, o resumo indicativo como deficiente. A análise dos dados feita aqui, aliás, provou que no corpus usado nesta pesquisa só havia resumos informativo-indicativos, e nenhum somente indicativo.

Os resumos informativo-indicativos, observados aqui na análise dos dados, não gozam integralmente da propriedade de representatividade. Eles não representam os originais, transmitindo consistentemente informações relevantes. Sua relação com os textos dos quais são sucedâneos é, em parte, referencial e dependente.

As deficiências específicas observadas nos resumos dos textos empíricos são apresentadas nos GRÁF. R3 e R4 abaixo.

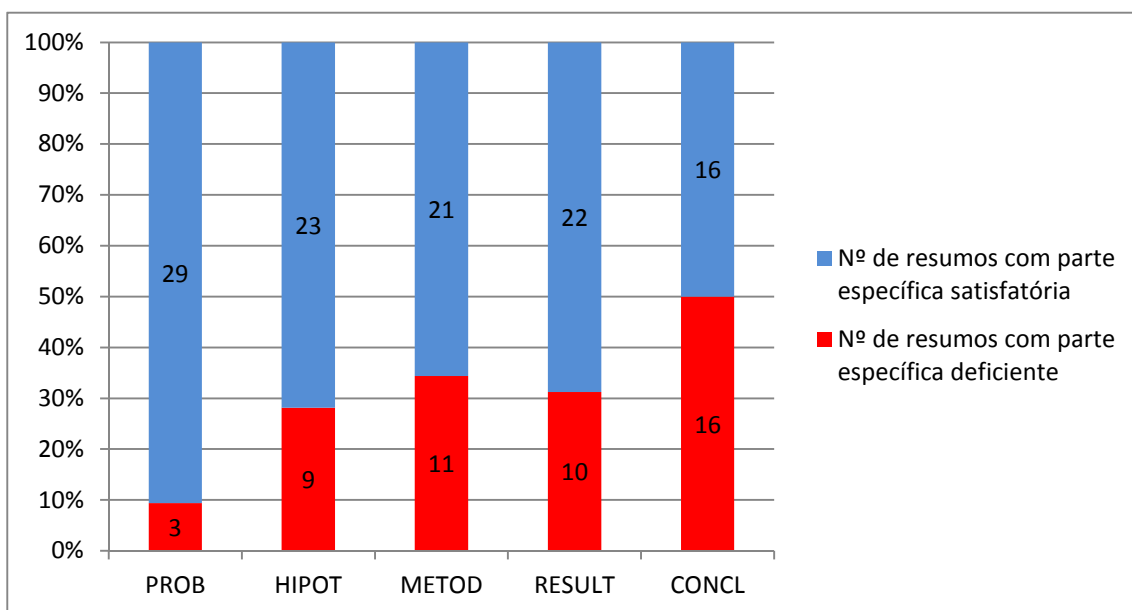


GRÁFICO R3 – Deficiências nos resumos do JD para artigos empíricos

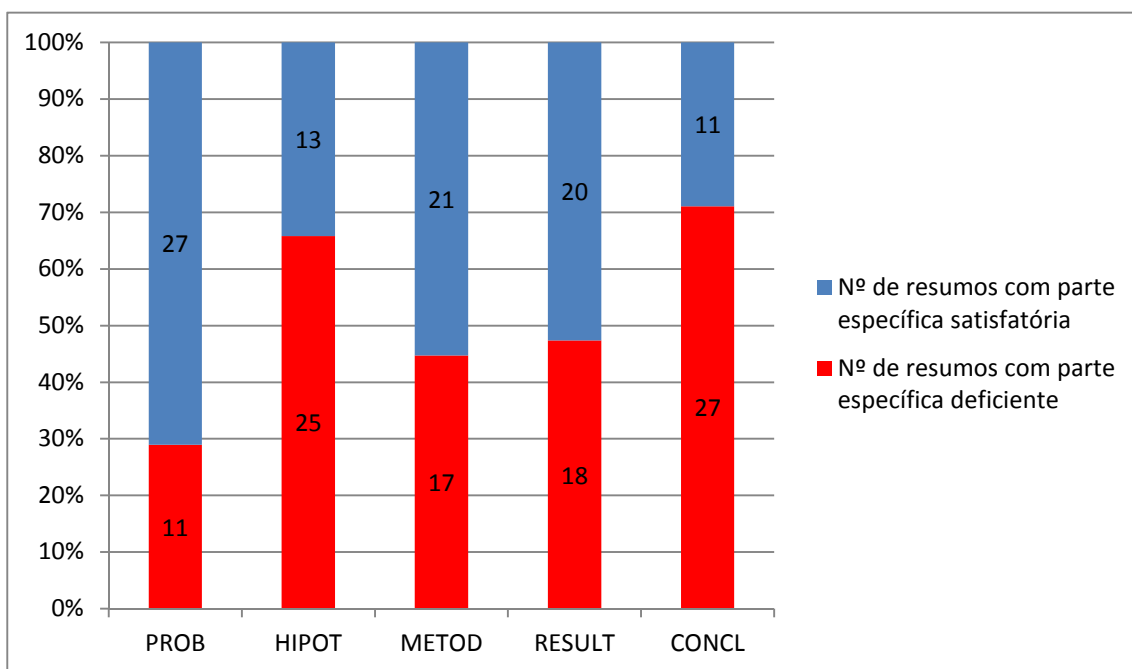


GRÁFICO R4 – Deficiências nos resumos do JASIST para artigos empíricos

O efeito da falta de padronização na atividade de seleção de conteúdos para a redação de resumo é visível nos GRÁF. R3 e R4, contudo, isso incide com maior prejuízo nos resumos informativos tradicionais que não oferecem nenhum recurso tópico para orientação da atividade de seleção, pois as orientações aos autores estão desligadas da redação do resumo propriamente dita. No caso dos resumos estruturados, parte importante dessa orientação é elemento preliminar e explícito constituinte do resumo e está no próprio espaço reservado à sua redação: propósito, metodologia/abordagem, resultados, conclusões e originalidade/valor.

As deficiências observadas nos diferentes itens da superestrutura, nos dois tipos de resumo, comprometeram a textualidade dos resumos. Acreditamos que os resultados numéricos que obtivemos oferecem apenas um indício do problema fundamental: parte dos resumos na comunicação científica não configuram textos dotados de coesão e coerência. Coesão e coerência não só tornam o resumo um produto documentário mais interpretativo como também singularizam sua similaridade com o texto original.

O problema da coesão e coerência textuais indicado aqui graficamente não admite uma análise puramente estatística. Por exemplo: um teste para a diferença entre as proporções de resumos informativos e estruturados deficientes não caracterizaria de que forma a textualidade dos resumos foi comprometida.

A detecção de segmentos textuais deficientes em um determinado resumo não dimensiona adequadamente a deficiência desse resumo enquanto texto. Seu impacto em nível global vai além de eventuais informações inconsistentes ou omitidas, na verdade, incide na própria capacidade de representação do produto documentário. Os números apresentados aqui servem como 'vestígios' de como a produção textual própria da atividade documentária pode ser afetada negativamente por falhas ligadas exatamente à textualidade.

A condensação textual continuará proporcionando uma representação sofrível para documentos tratados enquanto a seleção dos seus conteúdos for feita de forma assimétrica. Isso ficou bem caracterizado no tipo de deficiências encontradas no corpus da pesquisa: todas estavam associadas à omissão de conteúdo de um item da

superestrutura, e, a conteúdo de um item da superestrutura expresso em forma indicativa – o que, para fins de recuperação da informação, é uma forma de omissão.

Questões específicas ligadas à textualidade na produção de resumos também foram observadas por Narine, Yee, Einarson, e Ilersich (1991). Eles relataram que 66% dos resumos não forneciam dados para apoiar os resultados neles relatados. Esse tipo de falha, presente também nos resultados desta pesquisa, reduz o potencial comunicativo e a representação proporcionada pelo resumo. O resumo deve ser um texto independente em relação ao original e consistente nas relações que denota entre as diferentes partes de seu conteúdo.

Outro tipo de deficiência encontrada aqui e abordada por Salager-Meyer (1991) diz respeito à ‘dispersão conceitual’ (como resultados relatados em diferentes lugares do resumo). Como veremos adiante na reprodução integral de dois resumos considerados satisfatórios, a questão da dispersão conceitual é mais sutil e não uma simples deficiência. A sua ocorrência também está ligada a aspectos relacionados à textualidade, à representação coerente de ideias dentro de uma construção discursiva específica escolhida – e não só devido a uma inobservância de certa progressão de conteúdos que gerou a ‘deficiência’.

As noções de coesão e coerência são fundamentais para a textualidade de qualquer trabalho escrito. Entretanto, coesão textual não é *coerção* no sentido de que seja obedecida uma ordem fixa de apresentação das partes do conteúdo. O nexo entre as partes de um determinado conteúdo só vai depender de uma sequência convencional de apresentação se for construído assim. A coesão depende muito mais do modo como um determinado autor usa um estilo particular, da criatividade e clareza de sua composição. Isso pode ser alcançado de inúmeras maneiras. A ‘dispersão conceitual’ não se traduz necessariamente em falta de coesão ou em ineficiência na comunicação.

Algumas correlações puderam ser observadas no processo de análise. Tanto nos resumos estruturados quanto nos resumos informativos de artigos empíricos, as deficiências observadas na apresentação das hipóteses discutidas nos artigos originais (ou quando essas hipóteses não eram mencionadas no resumo) eram acompanhadas

de deficiências na apresentação dos resultados e conclusões desses artigos (ou de ausência de menção dos resultados e conclusões).

**NOS RESUMOS INFORMATIVOS DEFICIENTES (DE ARTIGOS EMPÍRICOS):**

12/35 ou 34,28% apresentavam deficiências na apresentação das hipóteses acompanhadas de deficiências na apresentação dos resultados;

19/35 ou 54,28% apresentavam deficiências na apresentação das hipóteses acompanhadas de deficiências na apresentação das conclusões;

9/35 ou 25,71% apresentavam deficiências na apresentação das hipóteses acompanhadas de deficiências na apresentação de resultados e conclusões.

**NOS RESUMOS ESTRUTURADOS DEFICIENTES (DE ARTIGOS EMPÍRICOS):**

5/18 ou 27,78% apresentavam deficiências na apresentação das hipóteses acompanhadas de deficiências na apresentação dos resultados;

7/18 ou 38,89% apresentavam deficiências na apresentação das hipóteses acompanhadas de deficiências na apresentação das conclusões;

5/18 ou 27,78% apresentavam deficiências na apresentação das hipóteses acompanhadas de deficiências na apresentação de resultados e conclusões.

As correlações apresentadas acima podem apontar para alguns padrões no processo de leitura e produção de resumos. Em ambos os tipos de resumo é alta a proporção (54,28% – informativos e 38,89% – estruturados) onde há incidência de falhas na apresentação de hipóteses e de conclusões concomitantemente.

Esse resultado tem a ver com o fato de que as pesquisas aplicadas do corpus desta pesquisa (como em geral acontece com pesquisas aplicadas em ciência da informação) lidam com problemas concretos e específicos, difíceis de serem generalizados por

hipóteses explicativas, e demasiado pontuais para serem antecipados por hipóteses preditivas. Isso é salientado por Hjørland (2000).

Muitas conclusões desses artigos são observações algo dispersas sobre aspectos particulares do problema estudado, o que torna difícil a ponderação e a escolha de só algumas delas para figurarem no resumo. Ou mesmo, se o autor considerar o conjunto de suas observações finais, como um todo, torna-se inviável sua condensação dentro do espaço previsto.



O texto dissertativo, por sua natureza e tradição filosófica, apresenta uma subjetividade muito incisiva – o que torna mais difícil sua condensação de uma forma sistemática. O item da superestrutura ‘argumentos’ aparece como o mais problemático. Ele pode abranger uma revisão de literatura, o histórico da tese a ser defendida, o levantamento de hipóteses a serem refutadas (o que pode ser feito imediatamente ou na conclusão), discussão dos métodos empregados pelos diferentes teóricos que abordaram o tema, etc. Tudo isso torna a condensação mais sujeita a erros e omissões.

No caso desse tipo textual, a conclusão também apresenta dificuldades para a redação do resumo porque costuma ser extensa e repartida, trazendo a termo muitos aspectos problematizados ao longo do texto. A conclusão, às vezes, não apresenta um caráter realmente original e inovador, trata apenas de corroborar opiniões e pontos de vista já consolidados na área. Isso faculta ao autor, quando redige o resumo, colocar ênfase naqueles aspectos do seu texto que são originais.

Os resultados relativos aos resumos dos artigos dissertativos aparecem nos GRÁF. R5 e R6 abaixo.

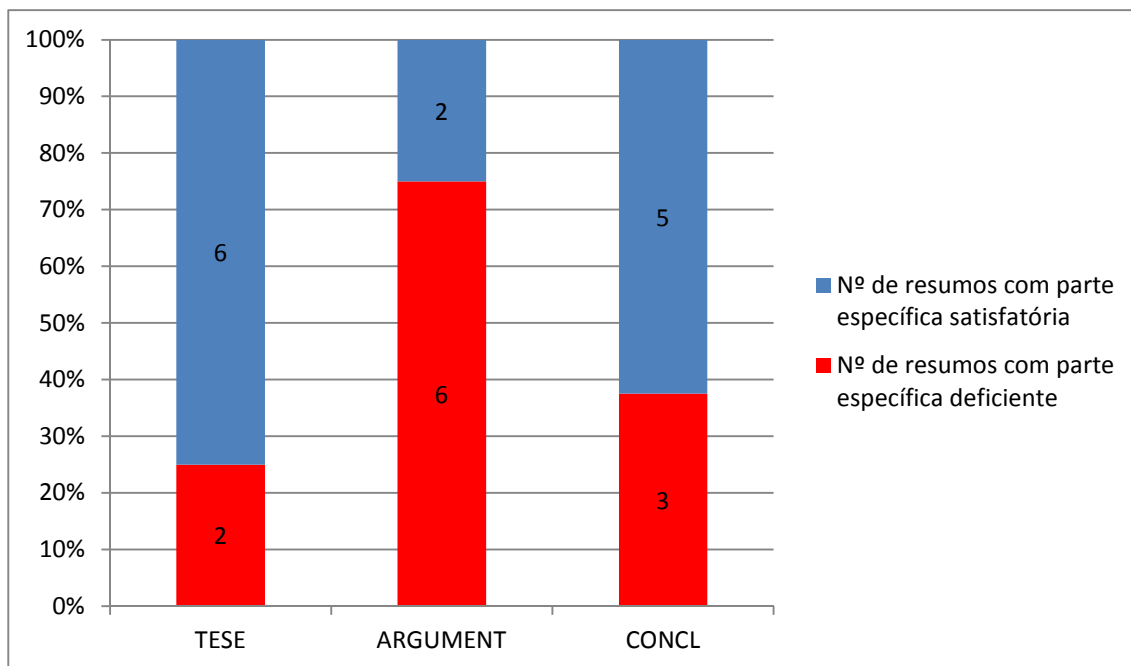


GRÁFICO R5 – Deficiências nos resumos do JD para artigos dissertativos

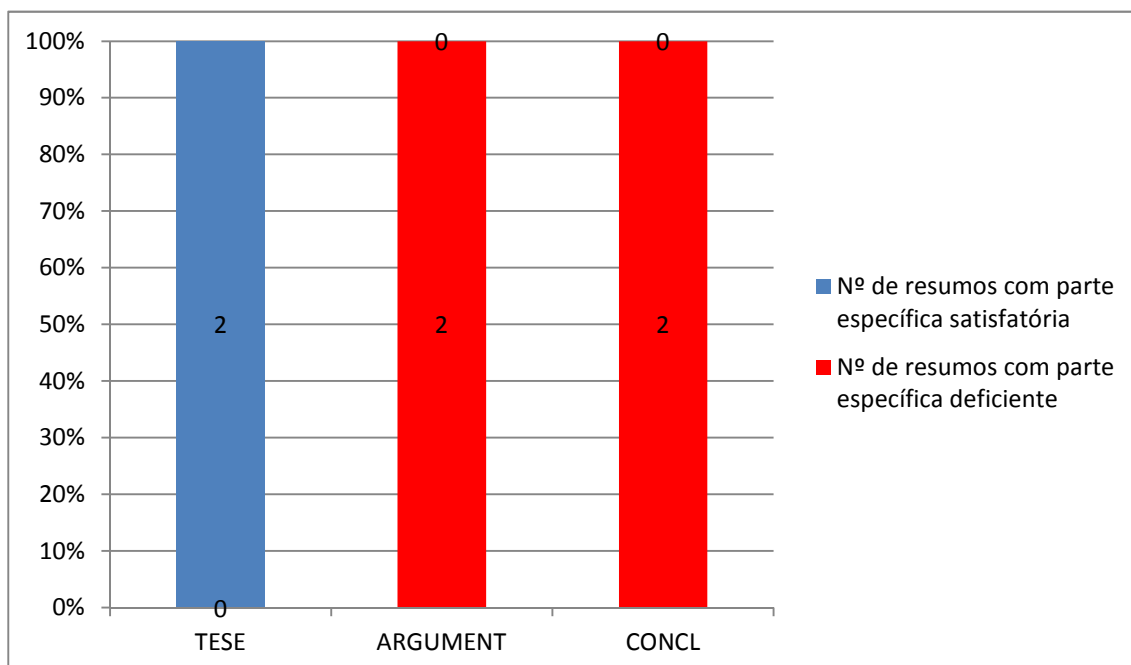


GRÁFICO R6 – Deficiências nos resumos do JASIST para artigos dissertativos

Os resumos de artigos expositivos apresentaram as seguintes deficiências:

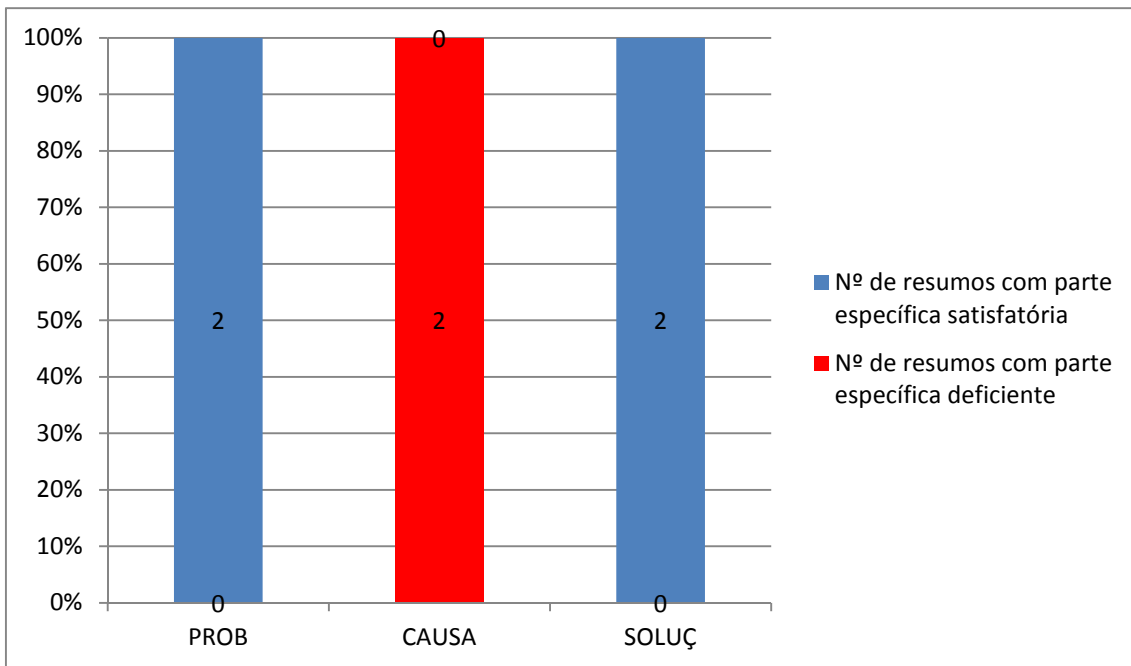


GRÁFICO R7 – Deficiências nos resumos do JD para artigos expositivos

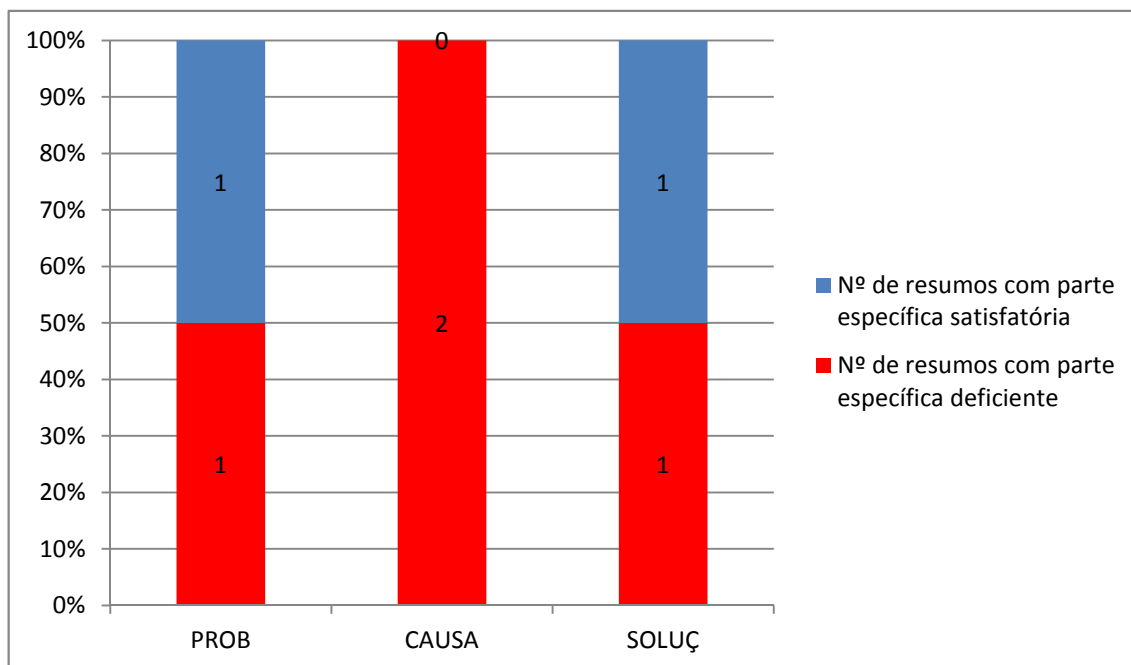


GRÁFICO R8 – Deficiências nos resumos do JASIST para artigos expositivos

Ambos os resumos dissertativos e expositivos apresentaram deficiências na representação do conteúdo relativo ao ‘desenvolvimento’ do texto do artigo original. Os resumos dissertativos falharam na apresentação de argumentos e os resumos expositivos falharam na apresentação das causas.

Os resumos informativos revelaram deficiências na apresentação de argumentos e conclusões em todas as duas ocorrências observadas de condensações de artigos dissertativos.

Os resumos estruturados (também para artigos dissertativos) revelaram deficiências na seguinte proporção:

1/8 ou 12,5% revelaram deficiências na apresentação de argumentos e conclusões;

5/8 ou 62,5% revelaram deficiências na apresentação apenas dos argumentos;

2/8 ou 25% revelaram deficiências na apresentação apenas das conclusões.



Com relação aos resumos expositivos, ambos os resumos informativos (dos dois que foram observados) revelaram deficiências na apresentação das causas, e, ambos os resumos estruturados também. Os resumos estruturados, no entanto, apresentaram corretamente (nos dois casos) o problema e a solução proposta.

O contato com resumos e artigos científicos em geral, revela um aspecto curioso da elaboração de resumos: muitos autores, no processo de condensação, ao invés de criarem uma representação do documento original criam algo como um material de divulgação (que, em recuperação da informação, é bem diferente de disseminação). Em seus resumos esses autores ‘mostram’ que escreveram um trabalho que deve ser consultado, e, de alguma forma, lhes escapa o fato de que o próprio resumo deve ser produzido como um item de consulta com efetivo valor informativo. Para isso, ele deve ter o ‘mesmo’ valor informativo do documento que ele representa.

Um exemplo de resumo informativo que atendeu aos critérios de similaridade e contiguidade textuais e que entendemos como satisfatório encontra-se na FIG. 8.

BALATSOUKAS, Panos; DEMIAN, Peter. Efeitos da granularidade dos resultados de busca no comportamento de julgamento de relevância de engenheiros: construindo sistemas para recuperação e entendimento do contexto. **Journal of the American Society for Information Science and Technology**, v. 61, n. 3, p. 453-468, 2010.

“Granularidade é um novo conceito para a apresentação de informações em interfaces de resultados busca em sistemas de recuperação de informação dirigidos por consulta hierárquica, de uma maneira que pode apoiar o entendimento e a exploração do contexto das informações recuperadas (por exemplo, destacando a sua posição na hierarquia granular e expondo sua relação com itens relacionados na hierarquia). Pouca pesquisa, no entanto, tem sido realizada sobre os efeitos da granularidade dos resultados de busca no comportamento de julgamento de relevância de engenheiros. Os engenheiros são usuários de informação altamente motivados, que estão particularmente interessados em compreender o contexto da informação recuperada. Portanto, é avançada a hipótese de que o projeto de sistemas que deem especial atenção à granularidade melhoraria o comportamento de julgamento de relevância de engenheiros. Para testar essa hipótese, um sistema protótipo foi desenvolvido e avaliado em termos do tempo necessário para os usuários encontrarem informações relevantes, a precisão do seu julgamento de relevância e a sua satisfação subjetiva. Para avaliar o protótipo, um estudo de usuário foi conduzido, onde os participantes foram solicitados a completar tarefas, preencher um questionário de satisfação, e participar de uma entrevista. Os resultados mostraram que os participantes tiveram melhor desempenho e ficaram mais satisfeitos quando o sistema protótipo apresentava apenas as informações relevantes no contexto. Embora este estudo apresente algumas novas descobertas sobre os efeitos de granularidade e do contexto no comportamento de julgamento de relevância do usuário, os resultados devem ser interpretados com cautela. Por exemplo, os participantes desta pesquisa foram recrutados por conveniência e realizaram um conjunto de tarefas simuladas em vez de tarefas reais. No entanto, sugestões para futuras pesquisas são apresentadas.”

**FIGURA 8 – Exemplo de resumo informativo/indicativo com informação integral conforme superestrutura**

No exemplo acima observa-se a expressão de conteúdo relevante dos diversos elementos da superestrutura do texto original. Uma solicitação de informação feita a um item documentário como esse é respondida com conteúdo ‘representativo’ (do original). Esse item documentário atende às expectativas que levanta. Isso acontece dada a similaridade que ele guarda com o texto original. Ele também guarda contiguidade com o original no nível da superestrutura primitiva e das relações que ela

articula. Há uma correspondência clara entre a sequência de base que ele determina e aquela do texto original.

O tema e o problema são apresentados com clareza nas três primeiras sentenças: o tema é o da 'granularidade' (aplicada a sistemas de recuperação da informação) cujas técnicas promovem entendimento e a exploração do contexto da informação recuperada. Um exemplo é dado para melhor compreensão do conceito.

O problema – que é delimitado com base na explicação do tema – envolve a especificidade do comportamento informacional dos engenheiros (particularmente interessados em entender o contexto da informação recuperada).

A hipótese é enunciada, tratando-se de uma hipótese preditiva ou ante-factum: um sistema verdadeiramente 'granular' deverá melhorar o julgamento de relevância por parte dos engenheiros.

A metodologia que testa a hipótese é explicada, delineando as características do 'protótipo' desenvolvido pelos autores. Aqui registra-se uma inconfidência: o dado de que são 12 estudantes de engenharia civil que participaram da pesquisa não foi especificado.

Os resultados são enunciados numa sentença (a saber, só informação relevante deve ser mostrada em contexto) e limitações do estudo são apontadas como dados da conclusão.

Um exemplo de resumo estruturado que também atendeu aos critérios de similaridade e (em grau avançado) de contiguidade textuais encontra-se na FIG. 9.

SAVOLAINEN, Reijo. Critérios de preferência de fonte no contexto de projetos do cotidiano: julgamentos de relevância feitos por indivíduos interessados na compra de imóveis. *Journal of Documentation*, v. 66, n. 1, p. 70-92, 2010.

**Propósito** – O propósito deste artigo é explorar como critérios de preferência de fonte são definidos no contexto de projetos do cotidiano que requerem busca de informação para problemas específicos. Mais especificamente, descobrir como aqueles que buscam por informação explicam seus critérios de preferência, caracterizando as forças e fraquezas percebidas nas diferentes fontes.

**Metodologia/abordagem** – A abordagem toma a forma de uma análise qualitativa de conteúdo de dados empíricos reunidos por entrevistas semiestruturadas com 16 interessados na compra de imóveis em 2008. Os critérios de preferência de fonte foram levantados por meio da construção do horizonte das fontes de informação.

**Resultados** – Fontes envolvendo rede de contatos foram mais favorecidas, seguidas por mídia impressa, fontes humanas e fontes organizacionais. O conteúdo da informação foi o critério primário de preferência de fonte. Disponibilidade da informação foi um critério bastante importante, enquanto que características do usuário, usabilidade da informação e fatores situacionais foram bastante marginais nesse respeito. Na definição de critérios de preferência, mais ênfase foi colocada nas forças do que nas fraquezas percebidas nas fontes. Qualidades positivas como “provê informação atualizada” foram referidas, particularmente enquanto se julgava a relevância das fontes que envolviam rede de contatos. Qualidades negativas como “informação desatualizada” foram primeiramente associadas com mídia impressa e fontes organizacionais.

**Limitações/implicações da pesquisa** – O estudo é exploratório, baseando-se numa amostra relativamente pequena recrutada através de um serviço disponível na web. Dessa forma, os resultados provêm não podem ser generalizados a potenciais interessados na compra de imóveis.

**Implicações práticas** – Interessados na compra de imóveis tendem a favorecer fontes e serviços disponíveis na web. Esses serviços deveriam prover os clientes com informação detalhada sobre o imóvel, incluindo fotos.

**Originalidade/valor** – O artigo especifica o modelo de julgamento de relevância usado pelo usuário no contexto da busca por informação no dia-a-dia.

**Tipo de artigo** – Pesquisa.

FIGURA 9 – Exemplo de resumo estruturado com informação integral conforme superestrutura

No exemplo acima, da mesma forma, podemos observar conteúdo dos diversos elementos da superestrutura do texto original. Sua formatação, envolvendo parágrafos e entretítulos, incrementa os efeitos que o uso da superestrutura textual traz para a compreensão dos diferentes conteúdos e como eles se relacionam.

O problema aparece enunciado no entretítulo 'propósito' e é bem especificado, com base na proposição acomodada no título do artigo (sendo o problema: como potenciais compradores de imóveis estabelecem critérios que definem sua preferência por uma fonte de informação sobre imóveis).

A hipótese explicativa ou post-factum aparece no entretítulo 'implicações práticas', o qual não é um campo obrigatório para a redação do resumo, mas é previsto o seu uso.

A metodologia aparece bem definida: entrevistas semi-estruturadas com 16 potenciais compradores de imóveis são analisadas qualitativamente. A definição dos indivíduos participantes do estudo fica clara e inequívoca.

Os resultados são bem caracterizados e as conclusões aparecem distribuídas na parte final do entretítulo 'resultados' e na parte final de 'implicações práticas' junto com a hipótese explicativa.

Esse resumo estruturado organiza a informação de modo que facilita a leitura e a análise. Ele permite uma avaliação de itens informacionais específicos (da superestrutura) de forma independente e também permite uma avaliação do conteúdo original de forma articulada.

## 5. Considerações Finais

Esta pesquisa comparou resumos tradicionais e resumos estruturados com os originais a partir dos quais foram preparados. Essa comparação se deu no nível da similaridade e contiguidade textuais. Depois foram comparados os dois tipos de resumo entre si, verificando qual deles comunicava mais informação relevante em relação à superestrutura textual primitiva usada no texto integral. Informação relevante, neste caso, foi definida como equivalente à similaridade e à contiguidade entre os dois textos – o artigo científico e seu resumo.

A avaliação da similaridade de conteúdo entre textos ainda não se encontra sistematizada e foi analisada segundo uma orientação subjetiva. Já a avaliação da relação de contiguidade entre textos pôde ser analisada significativamente e de forma padronizada pelo uso da superestrutura textual. Isso foi tanto mais eficaz quanto mais estruturalmente semelhantes os dois textos. A leitura do resumo feita independentemente do artigo que condensa também parece mais informativa a partir de um texto onde as ideias melhor se articulam. A contiguidade textual do resumo em relação ao artigo original replica melhor as relações entre ideias principais – desenvolvidas segundo uma cadeia discursiva esquematizada na superestrutura – facilitando a reprodução de um texto coerente e coeso.

Esse processo de avaliação se apoiou fundamentalmente nos tipos textuais estabelecidos por convenção cultural cujas características formam, principalmente, uma seqüência. Marcuschi (2002, p. 27) afirma que “quando se nomeia um certo texto como ‘narrativo’, ‘descritivo’ ou ‘argumentativo’ não se está nomeando o gênero e sim o predomínio de um tipo de sequência de base”. Essa sequência de base determina em grande medida o processo analítico que posteriormente vai orientar a síntese que o leitor faz do texto.

A noção de texto é basilar em todo o processo de comunicação científica. O resumo é o único produto documentário dotado de textualidade explícita. Essa conformidade entre produto documentário (resumo) e documento (artigo científico), entre texto integral e texto sucedâneo, deve ser convertida em vantagem para o processo comunicativo.

O resumo proporciona maior visibilidade e, principalmente, maior acessibilidade do documento que resume. Ele também garante precisão na recuperação da informação. No entanto, a revisão da literatura e os resultados apurados nesta pesquisa acusam falhas no processo de elaboração dos resumos que vão repercutir negativamente nas funções que eles cumprem. O processo de elaboração de resumos é um processo de produção textual e precisa atender às formalizações e conceituações próprias da produção de textos. Isso deve ser feito com a mesma austeridade e cuidado com que é produzido o artigo científico.

A prática documentária tem mostrado um descompromisso com a comunicação eficiente da informação dos artigos científicos de uma forma organizada em um nível profissional. O rigor procedimental que um pesquisador emprega na pesquisa científica é muito diferente do diletantismo com que ele executa a preparação do resumo e das palavras-chave que representarão sua pesquisa. A elaboração de resumos no processo de comunicação científica mais parece fruto de uma atividade livre e criativa de comunicação. Essa falha no processo não é corrigida pelo profissional da informação que na maioria dos casos nem tem autonomia para isso.

Nos resumos informativos era indistinto um padrão na apresentação dos conteúdos. Conforme mencionamos antes, em geral, os resumos apresentavam as ideias na mesma ordem em que essas apareciam nos originais, mas nos originais os autores não seguiram todos uma mesma ordem. Além disso, como a seleção das ideias para o resumo não era feita de forma padronizada, havia inconsistência no texto do resumo: às vezes não era apresentada a metodologia (ou ela era apenas indicada), outras vezes faltavam os resultados, etc. É bom salientar que a 'sequência de base' que um resumo deve obedecer é mais lógica do que restrita a uma progressão fixa de conteúdos.

Os resumos estruturados, nesse sentido, foram beneficiados pela estrutura de parágrafos e entretítulos compulsória (Propósito, Metodologia/Abordagem, Resultados, Conclusão) a ser observada pelos autores. Essa estrutura provou ajudar o autor a qualificar com mais atenção a informação que ele seleciona para representar um determinado item da superestrutura original.

A atividade documentária precisa avançar na elaboração de normas que regulem a composição dos *conteúdos* dos resumos. Precisam avançar os estudos sobre similaridade, coesão e coerência textuais.

Por outro lado, nada obsta os profissionais da informação de contribuírem para o avanço da área como um todo criativamente. A atividade documentária pode avançar, por exemplo, ao se envolver com a instrução a autores no que diz respeito à atribuição de títulos a trabalhos acadêmicos e científicos para publicação (DUARTE, 2007). Isso é importante para que se evitem falsas associações no momento da busca por informação.

A constatação de que o resumo de um artigo científico deve condensar conteúdo significativo das diferentes partes que compõem esse artigo científico parece mesmo um tanto óbvia e intuitiva. A questão, de imediato, é até um pouco desinteressante quando se pensa que hoje o acesso imediato ao texto completo é uma realidade e o resumo, em muitas circunstâncias, pode ser dispensado.

Com respeito à primeira constatação, o fato é que muitos resumos são preparados não como produtos documentários, mas algo como produtos de marketing. Produtos assim gerados conseguem a *apresentar* a informação mais do que *representá-la*. A informação assim veiculada pode até refletir aspectos importantes do processo comunicativo, mas não permite uma avaliação adequada da relevância da informação.

Quanto ao acesso direto ao texto completo, o resumo é indispensável para o exame ágil e eficiente de grandes volumes de informação. E é bom lembrar que o resumo é um item documentário, mas como texto independente ele também é subsidiário do processo de comunicação num nível intelectual.

Em comunicação científica, a leitura é sempre dirigida a uma finalidade bem determinada e é feita em profundidade de modo crítico e reflexivo. Inserido nesse quadro de aproximação entre indivíduo e conteúdo escrito está o resumo como instrumento de avaliação. O resumo atua reativamente na transição entre a leitura de reconhecimento – durante o processo de busca da informação – e a leitura reflexiva. Por meio do resumo, a atribuição de significado ao texto completo se processa como uma “catálise cognitiva”.

Por meio do resumo, as relações gerais entre os conceitos – comunicados pelo texto original – são delineadas. O leitor então entra em contato com o texto completo já dispondo de um arranjo conceitual sintético que é similar àquele mais amplo e complexo onde se introduz no momento da leitura reflexiva. O resumo, assim, coloca o leitor em uma situação estratégica. No entanto, essa propriedade estratégica do fenômeno da representação textual só é efetiva por meio da similaridade e contiguidade textuais. Os resumos analisados, de forma geral, não observaram criteriosamente a superestrutura textual para sua produção.

A atividade de condensação informacional, nos moldes como vem sendo feita hoje, não está provendo informação qualitativa ao leitor, mas quantitativa. Tal constatação não deixa de ser irônica. Em uma sociedade caracterizada como ‘sociedade da informação’ qualquer de seus membros já dispõe de informação abundante e não precisa de mais elementos de sobrecarga informacional. Não há sentido em um instrumento de avaliação – fruto de um laborioso exercício documentário – que proporcione apenas informação genérica e vaga e que não permita fazer distinção entre o que é de real interesse e o que apenas parece interessante.

A metodologia proposta por Kobashi (1997) representa um avanço no sentido de normalizar o processo de redação de resumos por meio de parâmetros claros, operacionalizando os conceitos de superestrutura textual para fins documentários. Trata-se de um avanço baseado em uma ideia simples e tão fundamental quanto generalizadamente desatendida. A proposta de Kobashi (1997) permite imprimir mais rigor, agilidade e objetividade ao processo documentário. Essa proposta auxilia a

pontuar e discriminar as informações a serem incluídas no resumo usando para isso um aspecto convencional dos textos na comunicação científica formal.

O formato do resumo estruturado é um avanço no mesmo sentido da proposta da Kobashi (1997). Sua estrutura compulsória adotada pelo JD provou trazer mais benefícios do que embaraços. O desempenho dos resumos estruturados na veiculação de informação relevante foi bem melhor do que o desempenho dos resumos meramente informativos. Isso aconteceu porque esse tipo de estruturação é inerente aos processos cognitivos humanos.

A esse respeito, Rendón Rojas (2005) esclarece que:

O problema sobre a objetividade da informação que surge porque ela é o resultado de uma atividade do sujeito, e, portanto, pode parecer subjetiva, é resolvido satisfatoriamente a partir de dois ângulos. Em primeiro lugar, a objetividade da informação é dada pelos dados que existem independentemente do sujeito [...] podemos diferenciá-los de outros dados, de alucinações. Em segundo lugar, **as estruturas que se utilizam para organização e significação do material empírico também têm o seu valor objetivo**, porque rejeitamos a ideia de que essas estruturas sejam inventadas arbitrariamente. As estruturas mencionadas são objetivas, porque são o resultado de um processo psicogenético que tem sua base na natureza biológica e psicológica do ser humano, ainda que, ao mesmo tempo, esse processo se ache imerso em contextos sócio-histórico-culturais determinados. Assim, cada sociedade institui seus esquemas para interpretar os símbolos com base nas relações que o ser humano estabelece com a natureza e com outros seres humanos no processo de produção material, espiritual e biológica. É evidente que as mencionadas relações e inter-relações são objetivas, pois permitem a própria existência da espécie humana e possibilitam o processo de imposição das vozes, o que U. [Umberto] Eco denomina "convenção cultural." (RENDÓN ROJAS, 2005, p. 53, o grifo é nosso).

A análise epistemológica tem grande potencial para promover o avanço do tratamento da informação na atividade documentária. Mas, a participação que a análise epistemológica formal tem hoje no tratamento da informação é desprezível e a elaboração teórica sobre os processos que ela envolve é escassa. A função que a análise epistemológica cumpre na interpretação da informação, entretanto, é crítica.

Sobre esse assunto, Laurence Heilprin (1989, p. 35) afirmou: “a falta de conhecimento da epistemologia é possivelmente a maior barreira ao avanço da ciência da e informação e biblioteconomia”.

Responder bem às questões ligadas ao tratamento da informação envolve princípios epistemológicos e implica em ver a recuperação da informação como um sistema cognitivo global, conforme observamos anteriormente (p. 34). Tal sistema cognitivo precisa distinguir melhor quais são aquelas estruturas responsáveis pela coesão e coerência de um dado conteúdo. Essas noções precisam ser melhor exploradas para o adiantamento do uso “infalível” de nossos processos cognitivos pois elas se ligam de forma importante à interpretação dos dados.

Segundo Varela (199-?) uma percepção mais apurada da atividade cognitiva – e, para nossos interesses, da recuperação da informação – é assim descrita:

Questão 1: O que é a cognição?

Resposta: A ação produtiva: o historial da união estrutural que emerge, que faz-emergir um mundo.

Questão 2: Como funciona?

Resposta: Por intermédio de uma rede de elementos interconectados, capaz de sofrer alterações estruturais ao longo de um historial não interrompido.

Questão 3: Como saber se um sistema cognitivo funciona de modo apropriado?

Resposta: Quando ele se une a um mundo de significados preexistente, em desenvolvimento contínuo (como é o caso dos descendentes de todas as espécies), ou quando forma um mundo

novo (como acontece na história da evolução). (VARELA, 199-?, p. 89).

A noção de coesão e coerência está vigorosamente implicada nas palavras de Varela (199-?). Um sistema cognitivo, segundo ele, só funciona adequadamente se for capaz de estabelecer algum tipo de coerência, de uniformidade entre seus 'elementos interconectados' de modo a reproduzir um sistema flexível e adaptável.

Conhecimento é refinamento da experiência, que transforma quantidade em qualidade, ensaio em destreza. Esse processo implica que o indivíduo se relacione coerentemente com o 'mundo de significados preexistente', ou, se envolva no aparecimento de um 'mundo' diferente.

O mundo real impõe ao ser humano controlar o processo de aprendizagem, padronizando-o, sintetizando-o e abreviando-o (GOODY, 1987, 1988); (LEROI-GOURHAN, 1984, 1990), ou, não aprender de forma nenhuma.

Gerar conhecimento envolve a superação dos obstáculos impostos ao processo de abstração. Segundo Oddone (2007), toda uma série de padrões documentários de organização e ordenamento do conhecimento permitiu que os processos de escrita e de leitura ampliassem suas chances de geração, recepção e reprodução do conhecimento, reduzindo as dificuldades de comunicação e aprendizagem. Conforme Oddone (2007, p. 109), técnicas e procedimentos como capítulos numerados em sequência, abertura de parágrafos, gráficos e tabelas, citações, etc. "disciplinam nosso pensamento e reduzem o caleidoscópico universo de agentes virtualmente cognitivos que nos rodeia, conformando a maneira como desenvolvemos e aplicamos nossos raciocínios".

Cabe à atividade documentária, por excelência, disciplinar a comunicação científica para que se realize uma comunicação eficiente.

"A ambição da complexidade é relatar articulações que são destruídas pelos cortes entre disciplinas, entre categorias cognitivas e entre tipos de conhecimento. De fato, a aspiração à complexidade

tende para o conhecimento multidimensional. Não se trata de dar todas as informações sobre um fenômeno estudado, mas de respeitar as suas diversas dimensões.” (MORIN, 2008, p. 138,).

“É, pois, necessário que toda a ciência se interrogue sobre as suas estruturas ideológicas e o seu enraizamento sociocultural. Aqui, damos-nos conta de que nos falta uma ciência capital, a ciência das coisas do espírito ou noologia, capaz de conceber como e em que condições culturais as ideias se agrupam, se encadeiam, se ajustam umas às outras, constituem sistemas que se auto-regulam, se autodefendem, se automultiplicam, se autoprogramam. Falta-nos uma sociologia do conhecimento científico que seja não só poderosa, mas também mais complexa que a ciência que examina.” (MORIN, 2008, p. 21).

A ciência da informação, mais do que uma ciência social aplicada, pode ser considerada uma sociologia do conhecimento aplicada. Uma ciência que se ocupa com a questão da complexidade envolvida na atividade de conhecer.

O problema da complexidade não é formular programas que os espíritos podem instalar no seu computador mental. [...] A complexidade faz apelo à estratégia. Só a estratégia pode ajudar a avançar no incerto e no aleatório. [...] A estratégia é a arte de utilizar as informações que surgem durante a ação, integrá-las, formular subitamente esquemas de ação e ser capaz de reunir o máximo de certezas para defrontar o incerto. (MORIN, 2008, p. 149).

## 6. Sugestões de Pesquisas

Acreditamos que as considerações feitas aqui a respeito de resumo indicativo são válidas e somam um conteúdo importante ao debate acerca da atividade de indexação e de redação de resumos.

A discussão realizada até aqui, além disso, nos habilita a fazer algumas sugestões. A elaboração de resumos, envolvendo tal carga de subjetividade como envolve, pode melhorar o desempenho de seu produto final no processo de recuperação da informação se conseguir fazer com que os resumos sejam tão discriminativos e singulares quanto possível.

Nesse sentido, a elaboração de resumos poderia ser realizada usando como expediente o levantamento prévio de uma amostra (para fins de comparação) de outros resumos de artigos científicos que abordam a mesma temática daquele artigo que se pretende resumir.

Esse artifício permite a quem vai redigir o resumo destacar, especificamente, os aspectos originais do novo artigo, tendo em vista aquilo que já foi publicado sobre o assunto e o modo como esses itens publicados foram tratados pelo trabalho de análise documentária.

Uma sugestão parecida foi feita por Haynes *et al.* (1990), embora esses autores tenham ido mais longe e pretendido que resumos do tipo estruturado ajudam leitores – profissionais de medicina – no exercício da clínica médica.

Outro ponto que merece mais estudo diz respeito à “vinculação de texto”. A vinculação de texto é uma técnica computacional que usa análises estatísticas e/ou sintáticas para identificar similaridades entre textos de diferentes documentos. Essa técnica pode ser adotada na produção automática de hiperlinks entre esses

documentos (LANCASTER, 2004, p. 327) oferecendo mais recursos para o aumento da precisão na recuperação da informação.

Estudos sobre coesão textual, observando uma abordagem interdisciplinar, parecem prementes para o avanço da atividade documentária. Em um texto que apresenta coesão, as ideias se ligam umas às outras em um fluxo natural para o senso comum. Esse processo precisa ser melhor compreendido e aproveitado em ciência da informação. Merece atenção especial a distinção entre coesão referencial, léxica e sequencial.

# Referências Bibliográficas

AD HOC WORKING GROUP FOR CRITICAL APPRAISAL OF THE MEDICAL LITERATURE. A Proposal for More informative Abstracts of Clinical Articles. **Annals of Internal Medicine**, v. 106, n. 4, p. 598-604, 1987.

ALVARENGA, Lídia. Alguns enunciados sobre comunicação e uso de fontes entre pesquisadores da área da educação. In: MUELLER, Suzana Pinheiro Machado; PASSOS, Edilenice Jovelina Lima (Org.). **Estudos avançados em ciência da informação**. Brasília, DF: Universidade de Brasília, 2000. p. 123-128.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6028: Resumos**. Rio de Janeiro, 2003.

ASTI VERA, Armando. **Metodologia da pesquisa científica**. Porto Alegre: Globo, 1980.

BACON, Francis. **Advancement of learning ; Novum organum ; New atlantis**. Chicago: Encyclopaedia Britannica, 1952.

BELL, Daniel. **O advento da sociedade pós-industrial: uma tentativa de previsão social**. São Paulo: Cultrix, 1978.

BERNAL, John Desmond. **Science in history**. London: Johnson's Court, 1954.

BERNAL, John Desmond. **Historia social de la ciencia**. Barcelona: Peninsula, 1976.

BORKO, Harold; BERNIER, Charles L. **Abstracting concepts and methods**. London: Academic Press, 1975.

BORKOWSKI, Casimir; MARTIN, J. Sperling. Structure, effectiveness and benefits of LEXtractor, an operational computer program for automatic extraction of case summaries and dispositions from court decisions. **Journal of the American Society for Information Science**, v. 26, n. 2, p. 94-102, 1975.

BOYLE, Robert; HUNTER, Michael Cyril William; DAVIS, Edward Bradford. **The works of Robert Boyle**. London: Pickering and Chatto, 1999.

BRIGGS, Asa; BURKE, Peter. **Uma história social da mídia: de Gutenberg à Internet**. Rio de Janeiro: Jorge Zahar, 2006.

BURKE, Peter. **Uma história social do conhecimento: de Gutenberg a Diderot**. Rio de Janeiro: Jorge Zahar, 2003.

BUSH, Vannevar. As we may think. **Atlantic Monthly**, v. 176, n. 1, p. 101-108, 1945. Disponível em: <<http://web.mit.edu/STS.035/www/PDFs/think.pdf>>. Acesso em: 5 out. 2010.

CAMPELLO, Bernadete S.; CAMPOS, Carlita Maria. **Fontes de informação especializada**: características e utilização. Belo Horizonte: UFMG, 1993.

CARNEIRO, Marília Vidigal. **Necessidades e demandas de informação dos técnicos do SEPLAN-MG**. 1982. Dissertação (Mestrado em Biblioteconomia) – Escola de Biblioteconomia, Universidade Federal de Minas Gerais, Belo Horizonte, 1982.

CASTELLS, Manuel. **O poder da identidade**. São Paulo: Paz e Terra, 2007. (Série: A era da informação: economia, sociedade e cultura).

CASTELLS, Manuel. **A sociedade em rede**. São Paulo: Paz e Terra, 2007. (Série: A era da informação: economia, sociedade e cultura).

CINTRA, Anna Maria Marques. Estratégias de leitura em documentação. In: SMIT, Johanna W. (Coord.). **Análise documentária**: a análise da síntese. Brasília: IBICT, 1987. p. 28-35.

CLEVELAND, Donald B.; CLEVELAND Ana D. **Introduction to indexing and abstracting**. Greenwood Village: Libraries Unlimited, 2001.

COPI, Irving M. **Introdução à lógica**. São Paulo: Mestre Jou, 1978.

CÔRTEZ, Pedro L. Considerações sobre a evolução da ciência e da comunicação científica. In: POBLACION, Dinah A.; WITTER, Geraldina P.; SILVA, José Fernando M. da (Org.) **Comunicação e produção científica**: contexto, indicadores e avaliação. São Paulo: Angellara, 2006. cap. 1, p. 33-56.

CRAVEN, Timothy C. Use of words and phrases from full text in abstracts. **Journal of Information Science**, v. 16, n. 6, p. 351-358, 1990.

CRAWFORD, S. Information needs and uses. **Annual review of information science and technology**, v. 13, p. 60-74, New Jersey: Learned Information, 1978.

CREMMNIS, Edward T. **The Art of Abstracting**. Arlington: IRP, 1996.

CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.

DANCY, Jonathan. Problems of epistemology. In: HONDERICH, Ted (Ed.). **The Oxford companion to philosophy**. Oxford: Oxford University, 1995.

DERTOUZOS, Michael L. **O que será**: como o novo mundo da informação transformará nossas vidas. São Paulo: Companhia das Letras, 1997.

DIAS, Eduardo Wense. Contexto Digital e Tratamento da Informação. **DataGramZero**, v. 2, n. 5. 2001. Disponível em: < [http://www.dgz.org.br/out01/Art\\_01.htm](http://www.dgz.org.br/out01/Art_01.htm)>. Acesso em: 5 out. 2010.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes. **Análise de assunto**: teoria e prática. Brasília: Thesaurus, 2007.

DRONBERGER, Gladys B.; KOWITZ, Gerald T. Abstract readability as a factor in information systems. **Journal of the American Society for Information Science**, v. 26, n. 2, p. 108-111, 1975.

DRUCKER, Peter Ferdinand. **The age of discontinuity**: guidelines to our changing society. New York: Harper and Row, 1969.

DRUCKER, Peter Ferdinand. **Administrando para o futuro**: os anos 90 e a virada do século. São Paulo: Pioneira, 1998.

DUARTE, Elizabeth A. **Comparação entre termos de indexação e palavras dos títulos dos artigos do periódico “cadernos de saúde pública” (2000/2005)**. 2007. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

EARL, Lois L. Experiments in automatic extracting and indexing. **Information Storage and Retrieval**, v. 6, p. 313-334, 1970.

FAYOL, M. Le résumé: un bilan provisoire des recherches de psychologie cognitive. In: CHAROLLES, M., PETITJEAN, A. **Le résumé de texte**: aspects linguistiques, sémiotiques, psycholinguistiques et automatiques. Paris: Klincksieck, 1991.

FERREIRA, Sueli Mara S. Pinto. **Estudo de necessidades de informação**: dos paradigmas tradicionais à abordagem sense-making. Porto Alegre: ABEBD, 1997. Versão eletrônica. (Documentos ABEBD). Disponível em: < <http://pt.scribd.com/doc/6747637/Estudo-de-Necessidades-de-InfornaCAo> >. Acesso em: 18 ago. 2010.

FERREIRA, Sueli Mara S. Pinto; MODESTO, F.; WEITZEL, Simone da R. Comunicação científica e o protocolo OAI: uma proposta na área de ciências da comunicação. **Comunicação e Sociedade**, Portugal, v. 6, p. 193-209, 2004.

FIDEL, Raya. Writing abstracts for free-text searching. **Journal of Documentation**, v. 42, n. 1, p. 11-21. 1986.

FOSKETT, Anthony Charles. **The subject approach to information**. London: 1998.

FROHMANN, Bernd. The role of the scientific paper in science information systems. In: **1998 Conference on the History and Heritage of Science Information Systems**, 1998, Medford, NJ. Proceedings of the... Medford: Information Today, 1999. p. 63-73.

FROOM, Paul; FROOM, Jack. Presentation deficiencies in structured medical abstracts. **Journal of Clinical Epidemiology**, v. 46, n. 7, p. 591-594, 1993.

GARDIN, Jean Claude. **La logique du plausible**: essais d'épistémologie pratique. Paris: Maison des Sciences de l'Homme, 1987. *apud* KOBASHI, Nair Yumiko. Resumos documentários: uma proposta metodológica. **Revista de Biblioteconomia de Brasília**, v. 21, n. 2, p. 201-210, 1997.

GARVEY, William D.; GRIFFITH, Belver C. Communication and information processing within scientific disciplines: empirical findings for psychology. **Information Storage and Retrieval**, v. 8, n. 3, p. 123-126, 1972.

GATES, Bill; MYHRVOLD, Nathan; RINEARSON, Peter. **A estrada do futuro**. São Paulo: Companhia das Letras, 1995.

GIASSON, Jocelyne; FRIAS, Maria José. **A compreensão na leitura**. Porto: Edições ASA, 1993.

GIDDENS, Anthony. **The nation state and violence**: a contemporary critique of historical materialism. Cambridge: Polity, 1985.

GIDDENS, Anthony. **Social theory and modern sociology**. Cambridge: Polity, 1987.

GIDDENS, Anthony. **Modernidade e identidade**. Rio de Janeiro: Zahar, 2002.

GINSPARG, Paul. First Steps toward Electronic Research Communication. In: DOWLER, Lawrence (Ed). **Gateways to knowledge**: the role of academic libraries in teaching, learning, and research. Boston: MIT, 1997.

GOODY, Jack. **A lógica da escrita e a organização da sociedade**. Lisboa: Editora 70, 1987.

GOODY, Jack. **Domesticação do pensamento selvagem**. Lisboa: Presença, 1988.

GREENE, Lewis Joel. O dilema do editor de uma revista biomédica: aceitar ou não aceitar. **Ciência da Informação**, v. 27, n. 2, p. 230-232, 1998.

GUIMARÃES, José Augusto C. Formas da informação jurídica: uma contribuição para sua abordagem temática. **Revista Brasileira de Biblioteconomia e Documentação**, v. 26, n. 1/2, p. 41-54, 1993.

HAGLER, Ronald. **The Bibliographic Record and Information Technology**. Chicago: American Library Association, 1991.

HAHS-VAUGHN, Debbie L.; ONWUEGBUZIE, Anthony J. Quality of abstracts in articles submitted to a scholarly journal: a mixed methods case study of the journal *Research in the Schools*. **Library and Information Science Research**, v. 32, n. 1, p. 53-61, 2010.

HARNAD, Stevan. **Categorical perception**: the groundwork of cognition. Cambridge: Cambridge University, 1990.

HARNAD, Stevan; *et al.* The access/impact problem and the green and gold roads to open access. **Serials review**, v. 30, n. 4, 2004. Disponível em: <<http://eprints.ecs.soton.ac.uk/10209/1/impact.html>>. Acesso em: 18 ago. 2010.

HARTER, Stephen P. Psychological relevance and information science. **Journal of the American Society for Information Science**, v. 43, p. 602-615, 1992.

HARTLEY, James. Three ways to improve the clarity of journal abstracts. **British Journal of Educational Psychology**, v. 64, n. 2, p. 331-343, 1994.

HARTLEY, James; SYDES, Matthew; BLURTON, Anthony. Obtaining information accurately and quickly: are structured abstracts more efficient? **Journal of Information Science**, v. 22, n. 5, p. 349-56, 1996.

HARTLEY, James; SYDES, Matthew. Are structured abstracts easier to read than traditional ones? **Journal of Research in Reading**, v. 20, n. 2, p. 122-136, 1997.

HARTLEY, James; BENJAMIN, M. An evaluation of structured abstracts in journals published by the British Psychological Society. **British Journal of Educational Psychology**, v. 68, p. 443-456, 1998.

HARTLEY, James. Is it appropriate to use structured abstracts in non-medical science journals? **Journal of Information Science**, v. 24, n. 5, p. 359-364, 1998.

HARTLEY, James. Applying ergonomics to Applied Ergonomics: using structured abstracts. **Applied Ergonomics**, v. 30, p. 535-541, 1999.

HARTLEY, James. Typographic settings for structured abstracts. **Journal of Technical Writing and Communication**, v. 30, n. 4, 355-365, 2000a.

HARTLEY, James. Are structured abstracts more or less accurate than traditional ones? A study in the psychological literature. **Journal of Information Science**, v. 26, n. 4, 273-277, 2000b.

HARTLEY, James. Do structured abstracts take more space? And does it matter? **Journal of Information Science**, v. 28, n. 5, 437-442, 2002.

HARTLEY, James. Improving the clarity of journal abstracts in Psychology: The Case for Structure. **Science Communication**, v. 24, n. 3, p. 366-379, 2003.

HARTLEY, James. Current findings from research on structured abstracts. **Journal of the Medical Library Association**, v. 92, n. 3, p. 368-371, 2004.

HARTLEY, James; BETTS, Lucy. The effects of spacing and titles on judgements of the effectiveness of structured abstracts. **Journal of the American Society for Information Science and Technology**, v. 58, n. 14, p. 2325-2340, 2007.

HARTLEY, James; BETTS, Lucy. Common weaknesses in traditional abstracts in the social sciences. **Journal of the American Society for Information Science and Technology**, v. 60, n. 10, p. 2010-2018, 2009.

HAYNES, R. B. *et al.* More informative abstracts revisited. **Annual internal medicine**, v. 113, n. 1, p. 69-76, 1990.

HAYNES, R. B. More informative abstracts: current status and evaluation. **Journal of clinical epidemiology**, v. 46, n. 7, p. 595-597, 1993.

HEILPRIN, Laurence B. Foundations of information science reexamined. **Annual Review of Information Science and Technology**, v. 24, p. 343-372, 1989.

HERSCHMAN, Arthur. The primary journal: past, present and future. **Journal of Chemical Documentation**, v. 10, n. 1, p. 37-40, 1970.

HJORLAND, Birger. Library and information science: practice, theory, and philosophical basis. **Information Processing and Management**, v. 36, p. 501-531, 2000.

HJORLAND, Birger. Toward a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. **Journal of the American Society for Information Science and Technology**, v. 52, p. 774-778, 2001.

HUCKIN, Thomas N. Surprise value in scientific discourse. In: **38th Meeting of the Conference on College Composition and Communication**, 1987, Atlanta, GA. Proceedings of the... Atlanta: Information Today, 1987.

HUTCHINS, John. Summarization: Some problems and methods, In: JONES, K. P. (Ed.). **Informatics 9: Meaning: The frontier of informatics**. London: Aslib, 1987. cap. 3, p. 151-173.

KAPLAN, Norman; STORER, Norman W. Scientific communication. In: SILLS, David L. **International encyclopedia of the social sciences**. New York: MacMillan & Free Press, 1968.

KING, Rosemary. A comparison of the readability of abstracts with their source documents. **Journal of the American Society for Information Science**, v. 27, n. 2, p. 118-121, 1976.

KOBASHI, Nair Yumiko. **A elaboração de informações documentárias: em busca de uma metodologia**. 1994. Tese (Doutorado em Comunicação) – Universidade de São Paulo, São Paulo, 1994.

KOBASHI, Nair Yumiko. Resumos documentários: uma proposta metodológica. **Revista de Biblioteconomia de Brasília**, v. 21, n. 2, p. 201-210, 1997.

KOCH, Ingedore Grunfeld Villaça. **A interação pela linguagem**. São Paulo: Contexto, 1992.

KUMAR, Krishan. **Da sociedade pós-industrial à pós-moderna**: novas teorias sobre o mundo contemporâneo. Rio de Janeiro: Jorge Zahar, 1997.

LANCASTER, Frederick. W. **Indexação e resumos**: teoria e prática. Brasília, DF: Brinquet de Lemos, 2004.

LAVILLE, Christian; DIONNE, Jean. **A construção do saber**: manual de metodologia da pesquisa em ciências humanas. Belo Horizonte: UFMG, 1999.

LAWRENCE, Steve. **Free online availability substantially increases a paper's impact**. Nature webdebates. 2001. Disponível em: <<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>>. Acesso em: 20 out. 2010.

LEONARD, L. E. **Inter-indexer consistency and retrieval effectiveness**: measurement of relationships. 1975. Thesis (PhD) – Universidade of Illinois, Urbana-Champaign, 1975. *apud* LANCASTER, Frederick. W. **Indexação e resumos**: teoria e prática. Brasília, DF: Brinquet de Lemos, 2004.

LEROI-GOURHAN, André. **Evolução e técnicas**. Lisboa: Setenta, 1984.

LEROI-GOURHAN, André. **O gesto e a palavra**. Lisboa: Setenta, 1990.

LUZ, Ana Cristina de Oliveira. **Condensação informacional**: aspectos e problemas em resumos informativos. 1992. Dissertação (Mestrado em Biblioteconomia e Documentação) – Universidade de Brasília, Brasília, 1992.

LUZ, Ana Cristina de Oliveira. Critérios para a elaboração de resumos. **Revista de Biblioteconomia de Brasília**, v. 20, n.1, p.27-40, 1996.

MACHLUP, Fritz. **International trade and the national income multiplier**. New York: Reprints of Economic Classics, 1961.

MACHLUP, Fritz; LEESON, Kenneth. **Information through the printed word**: the dissemination of scholarly, scientific, and intellectual knowledge. New York: 1978.

MAI, Jens-Erik. Analysis in indexing: document and domain centered approaches. **Information Processing and Management**, v. 41, p. 599-611, 2005.

MAIZELL, Robert E.; SMITH, Julian F.; SINGER, T. E. R. **Abstracting Scientific and technical literature**: an introductory guide and text for scientists, abstractors, and management. New York: Wiley-Interscience, 1971.

MARCUSCHI, Luiz Antônio. **Da fala para a escrita**: atividade de retextualização. São Paulo: Cortez, 2001.

MARTIN, James. **Manifesto**: presente e futuro da informática. São Paulo: Compucenter, 1983.

MÁXIMO, Antônio [Máximo] Ribeiro da Luz; ALVARENGA, Beatriz [Alvarenga] Álvares. **Curso de física**. São Paulo: Scipione, 1997.

MCINTOSH, N. Abstract information and structure: experience from an international paediatric meeting. **European science editing**, v. 23, n. 1, p. 3-6, 1997.

MCINTOSH, N. *et al.* Structure improves content and peer review of abstracts submitted to scientific meetings. **European science editing**, v. 25, n. 2, p. 43-47, 1999.

MEADOWS, Arthur J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MENEGHINI, R.; FONSECA, L. Índices alternativos de avaliação da produção científica em bioquímica no Brasil. **Ciência e cultura**, v. 42, n. 9, p. 629-645, 1990.

MERTON, Robert King. **A crítica da ciência**: sociologia e ideologia da ciência. Rio de Janeiro: Zahar, 1979.

MIKHAILOV, Aleksander Ivanovich; GILJAREVSKY, Rudhzero S. Estrutura e principais propriedades da informação científica (a propósito do escopo da informática) In: FOSKETT, Douglas J. **Ciência da informação ou informática?** Rio de Janeiro: Calunga, 1980.

MIRANDA, Antônio. **Ciência da informação**: teoria e metodologia de uma área em expansão. Brasília: Thesaurus, 2003.

MIRANDA, Dely Bezerra de. O periódico científico como veículo de comunicação: uma revisão da literatura. **Ciência da informação**, v. 25, n. 3, p. 375-382, 1996.

MOLES, Abraham A. **A criação científica**. São Paulo: Perspectiva: USP, 1971.

MOLINA, Maria P. Documentary abstracting: toward a methodological model. **Journal of the American Society for Information Science**, v. 46, n. 3, p. 225-234, 1995.

MOREIRO GONZÁLEZ, José A. **Aplicación de las ciencias del texto al resumem documental**. Madrid: Universidad Carlos III, 1993.

MORIN, Edgar. **Ciência com consciência**. Rio de Janeiro: Bertrand Brasil, 2008.

MUELLER, Suzana Pinheiro Machado. O impacto das tecnologias da informação na geração do artigo científico: tópicos para estudo. **Ciência da Informação**, v. 23, n. 3, p. 309-317, 1994.

MUELLER, Suzana Pinheiro Machado. Realidade e controvérsia das publicações eletrônicas: o periódico científico. **Revista Biblioteconomia de Brasília**, v. 21, n. 1, p. 109-130, 1997.

MUELLER, Suzana Pinheiro Machado. O círculo vicioso que prende as revistas nacionais. **Datagramazero**, v. 0, n. 0, texto em html, 1999.

MUELLER, Suzana Pinheiro Machado. A ciência, o sistema de comunicação científica e a literatura científica. In: CAMPELLO, Bernadete Santos; CENDÓN, Beatriz Valadares; KREMER, Jeannette Marguerite (Org.) **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: UFMG, 2000. cap. 1, p. 21-34.

MUELLER, Suzana Pinheiro Machado. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v. 35, n. 2, p. 27-38, 2006. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/view/826/668> Acesso em: 5 jun. 2010.

MUELLER, Suzana Pinheiro Machado. Literatura científica, comunicação científica e ciência da informação. In: TOUTAIN, Lídia Maria B. B. (Org.) **Para entender a ciência da informação**. Salvador: UFBA, 2007. p. 125-144.

NAISBITT, John. **Megatendências**: as dez grandes transformações que estão ocorrendo na sociedade moderna. São Paulo: Abril Cultural, 1982.

NARINE, Lutchmie; YEE, Deanna S.; EINARSON, Thomas R.; ILERSICH, A. Lane. Quality of abstracts of original research articles in CMAJ in 1989. **The Canadian Medical Association Journal**, v. 144, n. 4, p. 449-453, 1991.

NEGROPONTE, Nicholas. **Being digital**. New York: Vintage Books, 1996.

ODDONE, Nanci E. Revisitando a “epistemologia social”: esboço de uma ecologia sociotécnica do trabalho intelectual. **Ciência da Informação**, v. 36, n. 1, p. 108-123, 2007.

OHIRA, Maria Lourdes Blatt; SOMBRIO, Márcia Luiza Lonzetti Nunes; PRADO, Noêmia Schoffen. Periódicos brasileiros especializados em biblioteconomia e ciência da informação: evolução. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 5, n. 10, 2000. Disponível em: <[http://www.encontrosbibli.ufsc.br/Edicao\\_10/lurdinha.htm](http://www.encontrosbibli.ufsc.br/Edicao_10/lurdinha.htm)>. Acesso em: 24 maio 2010.

OLIVEIRA, Marlene. A pesquisa científica na ciência da informação: análise da pesquisa financiada pelo CNPq. **Perspectivas em Ciência da Informação**, v. 6, n. 2, p. 143-156, 2001.

OMNÈS, Roland. **Filosofia da ciência contemporânea**. São Paulo: UNESP, 1996.

PÁLSDÓTTIR, Ágústa. The connection between purposive information seeking and information encountering: a study of icelander's health and lifestyle information seeking. **Journal of Documentation**, v. 66, n. 2, p. 224-244, 2010.

PINTO, María; GÁLVEZ, Carmen. Paradigms for abstracting systems. **Journal of Information Science**, v. 25, n. 5, p. 365-380, 1999.

PITKIN, Roy M.; BRANAGAN, Mary Ann. Can the Accuracy of Abstracts Be Improved by Providing Specific Instructions: A Randomized Controlled Trial. **Journal of the American Medical Association**, v. 280, n. 3, p. 267-269, 1998.

PITKIN, Roy M.; BRANAGAN, Mary Ann; BURMEISTER, Leon F. Accuracy of data in abstracts of published research articles. **Journal of the American Medical Association**, v. 281, n. 12, p. 1110-1111, 1999.

PITKIN, Roy M.; BRANAGAN, Mary Ann; BURMEISTER, Leon F. Effectiveness of a Journal Intervention to Improve Abstract Quality. **Journal of the American Medical Association**, v. 283, n. 4, p. 481-487, 2000.

POBLACIÓN, Dinah A.; OLIVEIRA, Marlene. Input e Output: Insumos para o desenvolvimento da pesquisa. In: POBLACION, Dinah A.; WITTER, Geraldina P.; SILVA, José Fernando M. da (Org.) **Comunicação e produção científica: contexto, indicadores e avaliação**. São Paulo: Angellara, 2006. cap. 2, p. 57-66.

POLANYI, Michael. **Personal knowledge: towards a post-critical philosophy**. Chicago: University of Chicago Press, 1962.

PORAT, Marc Uri. **The Information economy: definition and measurement**. Washington: Superintendent of Documents, U.S. Government Printing Office, 1977.

PRICE, Derek J. de Solla. **Little science, big science**. New York: Columbia University Press, 1971.

PRICE, Derek J. de Solla. **O desenvolvimento da ciência: análise histórica, filosófica, sociológica e econômica**. Rio de Janeiro: Livros Técnicos e Científicos, 1976.

RENDÓN ROJAS, Miguel Angel. Relación entre los conceptos: Información, conocimiento y valor: Semejanzas y diferencias. **Ciência da Informação**, v. 34, n. 2, p. 52-61, 2005.

ROBREDO, Jaime. Filosofia da ciência da informação ou ciência da informação e filosofia. In: TOUTAIN, Lídia Maria B. B. (Org.) **Para entender a ciência da informação**. Salvador: UFBA, 2007.

ROSSI, Paolo. **A ciência e a filosofia dos modernos**: aspectos da revolução científica. São Paulo: UNESP, 1992.

ROWLEY, Jennifer. E. **Abstracting and indexing**. London: 1988.

SALAGER-MEYER, Françoise. Medical English abstracts: How well are they structured? **Journal of the American Society for Information Science**, v. 42, n. 7, p. 528-531, 1991.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 41-62, 1996.

SAUPERL, Alenka; KLASINC, Janko; LUZAR, Simona. Components of Abstracts: Logical Structure of Scholarly Abstracts in Pharmacology, Sociology, and Linguistics and Literature. **Journal of the American society for information science and technology**, v. 59, n. 9, p. 1420-1432, 2008.

SERAFINI, Maria Teresa. **Como escrever textos**. Rio de Janeiro: 1987.

SHERA, Jesse Hank; FOSKETT, Douglas J. **Documentation and the organization of knowledge**. Hamden: 1966.

SHERA, Jesse Hank. **The foundations of education for librarianship**. New York: Becker and Hayes, 1972.

SILVA, Edna Lúcia *et al.* Panorama da Pesquisa em Ciência da Informação. **Informação e sociedade**: Estudos, v. 16, n. 1, p. 159-177, 2006.

SLIFE, B. D.; WILLIAMS, R. N. **What's behind the research?** Discovering hidden assumptions in the behavioral sciences. London: Sage Publications, 1995.

SOUFFLET, Edmond; LOISEAU, Maryvonne. **Le résumé et la contraction de texte**: aux concours et examens: (grandes écoles): méthode et épreuves. Paris: Vuibert, 1971.

SPERBER, Dan; WILSON, Deirdre. **Relevância**: comunicação e cognição. Lisboa: Fundação Calouste gulbenkian, 2001.

STEHR, Nico. **Knowledge societies**. New York: Sage, 1994.

STUBBS, Edgard A.; MANGIATERRA, Norma E.; MARTINEZ, Ana M. Internal Quality Audit of Indexing: A New Application of Interindexer Consistency. **Cataloging and Classification Quarterly**, v. 28, n. 4, p. 53-69, 2000.

TADDIO, A.; PAIN, T.; FASSOS F. F.; BOON, H.; ILERSICH, A. L.; EINARSON, T. R. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. **The Canadian Medical Association Journal**, v. 150, n. 10, p. 1611-1615, 1994.

TENOPIR, Carol; JACSÓ, Péter. Quality of abstracts. **Online**, n. 5, p. 44-54, 1993.

TIBBO, Helen. Abstracting across the disciplines: a content analysis of abstracts from the natural sciences, the social sciences, and the humanities with implications for abstracting standards and online information retrieval. **Library and Information Science Research**, n. 14, p. 31-56, 1992.

TOFFLER, Alvin. **A terceira onda**. Rio de Janeiro: Record, 1980.

TOFFLER, Alvin. **O choque do futuro**. Rio de Janeiro: Record, 1994.

TRAWINSKI, Bogdan. A methodology for writing problem structured abstracts. **Information processing and management**, v. 25, n. 6, p. 693-702, 1989.

VAN DIJK, Teun Adrianus; KINTSCH, Walter. **Strategies of discourse comprehension**. Orlando: Academic Press, 1983.

VARELA, Francisco J. **Conhecer**: as ciências cognitivas, tendências e perspectivas. Lisboa: Inst. Piaget, [199-?].

VINSONHALER, John F. Some behavioral indices of the validity of document abstracts. **Information Storage and Retrieval**, v. 3, n. 1, p. 1-11, 1966.

WEBSTER, Frank. **Theories of the information society**. New York: Routledge, 2006.

WEIL, B. H.; ZAREMBER, I.; OWEN, H. Technical-abstracting fundamentals: writing principles and practices. **Journal of Chemical Documentation**, v. 3, n. 3, p. 125-132, 1963.

WEINBERG, Bella Hass. Why indexing fails the researcher. **The Indexer**. v. 16, n. 1, p. 3-6, 1988.

WEISMAN, Herman M. **Information systems, services and centers**. New York: Wiley-Becker-Hayes, 1972.

WEITZEL, Simone da Rocha. Fluxo da informação científica. In: POBLACION, Dinah A.; WITTER, Geraldina P.; SILVA, José Fernando M. da (Org.) **Comunicação e produção científica**: contexto, indicadores e avaliação. São Paulo: Angellara, 2006. cap. 3, p. 81-114.

WERSIG, Gernot; NEVELING, Ulrich. The phenomena of interest to information science. **The Information Scientist**, v. 9, n. 4, p. 127-140, 1975.

WHITEHEAD, Alfred North; RUSSELL, Bertrand. **Principia mathematica**. Cambridge: University of Cambridge, 1950.

ZANDONADE, Tarcísio. Social epistemology: from Jesse Shera to Steve Fuller. **Library Trends**, v. 52, n. 4, p. 810-832, 2004.

ZIMAN, John. **Conhecimento público**. Campinas: Papyrus, 1979.

## Apêndice A

### **Consulta a pesquisadores em ciência da informação e engenharia**

Como forma de enriquecer a pesquisa documental conduzida aqui, pedimos que professores e pesquisadores em ciência da informação e engenharia avaliassem a) uma característica do resumo estruturado, e, b) uma sugestão que propõe alterações na forma como é apresentado o resumo informativo tradicional.

Nossa intenção com essa consulta a profissionais envolvidos na comunicação científica era de obter uma apreciação desses indivíduos também como usuários-finais. A escolha de professores em Ciência da Informação para a consulta se deveu ao fato de que os periódicos analisados aqui eram específicos dessa área. Como contraponto, consultamos também professores em Engenharia. Essa escolha foi determinada por conveniência do pesquisador e porque as observações colhidas formam um contraponto tão bom como seria o caso com qualquer outro grupo de professores. Além disso, procuramos oferecer – para consideração desses profissionais – perspectivas que constituíssem iniciativas diferentes de promover avanços no processo de comunicação científica.

Dez pesquisadores das áreas de Ciência da Informação e Engenharia foram inquiridos. Primeiro, pedimos aos professores que se pronunciassem em relação à utilidade da enunciação formal da *originalidade e valor* de determinado artigo científico junto ao resumo (conforme dita a política do JD para todo resumo em suas publicações).

## **Comentários a respeito da utilidade da enunciação formal, feita no resumo, da originalidade e valor do artigo publicado.**

### **Comentário 1 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“Eu acho interessante, mas, eu acho difícil que o autor tenha uma percepção clara/imparcial da própria contribuição... como é um atributo de valor... Porque como se trata de “originalidade e valor” isso pode ser inflado pelo autor. Esse item, no resumo, eu acho que foge ao propósito. O resumo é um ‘conteúdo informacional condensado’, de modo que, a partir do resumo, o leitor é capaz de avaliar: esse artigo me interessa, esse artigo não me interessa.”

### **Comentário 2 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“Eu acho que isso é pertinente, uma vez que facilita a consulta para o leitor – seja ele especialista ou leitor leigo – a compreensão da originalidade do trabalho, sobretudo na área da pesquisa. Quando se trata de um leitor que também é pesquisador torna-se, então, mais fácil, a partir de um resumo com esses dados, obter uma visão do que o próprio autor considera original no seu trabalho de pesquisa.”

### **Comentário 3 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“É legal a pessoa dizer qual o diferencial do trabalho dela, ou, pelo menos, o que ela entende que é o diferencial no trabalho dela. Eu fico imaginando se o diferencial que o autor enxerga é o diferencial para quem está lendo também. É uma diferença sutil.

Eu vejo isso como uma tentativa de oferecer para a pessoa que está lendo o resumo, uma informação extra que talvez vá fazer com que ela se interesse em ler com mais vagar esse artigo. Se é útil, de uma forma bem ampla, eu não poderia dizer, mas eu acho que é uma contribuição que mal não faz. Eu não vejo problema nenhum em que se peça isso. Se a pessoa não for capaz de dizer o que é relevante no trabalho dela, ela não deveria nem ter publicado. Se eu vou achar relevante ou não, se eu vou ler ou não, pelo menos, isso já me oferece um dado extra... Então, eu acho que sim, pode ser útil.”

### **Comentário 4 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“O resumo veio para poder informar o leitor sobre o que ele vai encontrar no trabalho, isso é obrigatório. Mais do que importante, é uma necessidade. Isso que você introduz de ferramental de tecnologia, nós ensinamos nossos alunos a fazerem desse mesmo jeito. O que você está ganhando quando segue essa metodologia? [do resumo estruturado]

Esse formato é burocrático. Ele pode ser educativo, porque é isso que nós fazemos com nossos alunos. Se você vai escrever um projeto, você precisa dizer qual o seu objetivo, escopo, metodologia, o que você espera alcançar. Um resumo desse jeito é necessário para treinar. Ele explica, mas, ele não é elucidativo para quem vai ler. A gente já espera encontrar essas informações.

O leitor pode se aproximar do resumo com a seguinte preocupação: ‘Qual a originalidade?’, mas, a metodologia é um passo essencial. A metodologia, por exemplo, pode ser feita aqui ou acolá, mas, o problema do artigo é tratável por essa

metodologia? Você não pode chamar a atenção para algumas coisas, porque o resumo é um todo encadeado, mas, isso aqui é bom para o iniciante... educativo ele é. Mas, o resumo estruturado pode ser enganoso, apontando para determinados pontos sem dar uma ideia geral do texto.”

#### **Comentário 5 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“É uma questão muito complicada. Pode ser interessante se for condicionado a... porque o próprio autor é quem vai colocar qual é o valor do artigo dele. O autor, no entanto, está sujeito a extrapolar a contribuição efetiva que fez. Em tese, eu acho que é interessante, mas condicionado a revisões.”

#### **Comentário 6 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“Nesse tipo de resumo (estruturado)... eu não sei. É interessante destacar a originalidade porque a gente vê tanta coisa igual. Eu, particularmente, prefiro o resumo tradicional. Ele me fala tudo, [ ] e a partir dele você já sabe se ele é original. Mas, respondendo a sua pergunta, eu acho que seria interessante, o autor tem que falar porque seu artigo é original, o que ele acha que é original. Eu gosto do resumo tradicional porque, se o autor sabe o que está fazendo, ele põe tudo no resumo e você faz uma leitura uniforme, contínua, e você sabe se o artigo te interessa ou não. Se o autor não sabe fazer, você lê, e o autor não falou nada ali e você continua o processo de busca. O autor perde um cliente. O resumo é muito importante, mas, tem gente que não liga porque você muitas vezes lê o resumo e não sabe o que o autor fez, você não se interessa pelo artigo e o autor perde um cliente.”

### **Comentário 7 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“A gente normalmente coloca [qual a relevância do trabalho]... você diz, afinal de contas, para quê você está fazendo aquele trabalho. Eu acho que tem que constar nos dois tipos de resumo. Se muda o tipo de resumo, tem que constar a mesma coisa aqui, então, você diz: “o propósito desse trabalho é...” e o leitor vai saber o que você está fazendo, qual é a distinção daquele trabalho em relação aos demais.”

### **Comentário 8 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“O artigo deve buscar a originalidade, isso é um aspecto inerente à pesquisa. Eu acho positiva essa iniciativa, eu acho interessante. É uma proposta diferente...”

### **Comentário 9 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“Eu acho interessante... eu não faria um resumo assim, para ser sincero. Eu acho interessante porque é muito mais rápida a informação e normalmente em um resumo a gente quer informação direta. Esse resumo nos permite avaliar se um trabalho nos interessa. Eu acho que algumas pessoas vão estranhar, vão ter dificuldade em falar da originalidade e valor do seu trabalho, às vezes as pessoas nem se preocupam muito com esse item (até antes mesmo de fazer a pesquisa) e isso vai começar a chamar a atenção. É algo fundamental para os trabalhos científicos, que eles tenham um valor claro, isso tem que ficar muito claro para o pesquisador e para quem está lendo o trabalho. Eu acho interessante.”

## **Comentário 10 sobre o item “Originalidade/valor” obrigatório no resumo estruturado do JD**

“Eu não vejo o resumo estruturado como uma coisa inovadora. É colocar a informação numa fôrma. Eu acho interessante o seguinte, na verdade, isso aqui tem que figurar na introdução de um Trabalho de Graduação. Eu repito isso para os meus alunos. Quando eu avalio um pôster, eu vejo que vários aspectos do trabalho são pormenorizados, mas não se escreve uma linha dizendo por que o trabalho é importante para o aluno, ou importante de uma forma geral. Eu posso melhor avaliar um trabalho se eu sei aquilo que é importante para o autor. A minha leitura dessa proposta é a seguinte: muitos artigos vão destacar sua originalidade e valor e, na verdade, é tudo baboseira porque o artigo em si não tem originalidade mesmo. Isso não faz parte de um resumo analítico. O resumo deveria ser: objetivo, materiais e métodos, resultado e conclusão, e deveria dispensar a leitura do artigo. Eu acho que é uma inovação você colocar uma justificativa, você dizer porque o seu trabalho é importante. A originalidade de um trabalho, em geral, é um detalhe, é 1% do trabalho: “eu fiz um trabalho semelhante ao de fulano, mas utilizei o método ‘b’”. É um detalhe, mas alguém tinha que provar isso.”

## **Comentários a respeito da utilidade de indicação numérica de parágrafo, relacionando a informação do texto original com a informação encontrada no resumo.**

Perguntamos, também, se seria útil oferecer no corpo do texto de um resumo informativo tradicional *indicação numérica do(s) parágrafo(s)* onde aquela informação aparecia discutida em detalhe no texto original. Nesse caso, os parágrafos do texto original apareceriam numerados. Suas impressões foram essas:

**Comentário 1 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“Acho interessante, mas, eu vejo a norma técnica e vejo tantos resumos mal-feitos, que eu acho que, primeiro, devemos nos preocupar com que se faça um bom resumo, para depois acrescentar alguma inovação. Eu me interesso muito pelo assunto e vejo muitos resumos errados. O autor, muitas vezes, faz o resumo desconhecendo as normas, desconhecendo o que é o resumo. Certos resumos não reportam os resultados, outros não falam da metodologia, etc. Você lê o resumo e não sabe absolutamente qual é o objeto do trabalho.”

**Comentário 2 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“Não acredito que isso seja pertinente, porque o resumo, assim, passa a ter uma dimensão ampliada para o pesquisador, mais do que, talvez, um resumo estruturado. Eu acredito que isso possa trazer um desconforto para o pesquisador: ter que citar qual o parágrafo onde ele reproduz uma informação.

Se o pesquisador acredita que citar quais são os parágrafos que sustentam trechos do resumo é relevante, não me oponho, mas, para isso se configurar em norma técnica, eu já não sou a favor. Quando o pesquisador escreve o resumo, ele já está sendo fiel às suas concepções, porque isso já é uma exigência primordial, a ideia básica do resumo é essa. O uso dessa marcação de parágrafos eu deixaria em aberto, ficaria a critério do autor. Mas eu não acredito que isso é um ponto de extrema relevância para constar num resumo. Remeter àqueles parágrafos que tornam aquele resumo mais confiável, não parece pertinente... talvez na área do direito seja pertinente, onde há uma necessidade de que se remeta à lei, ... ou nas *hard sciences*.”

**Comentário 3 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“Com certeza é útil, porque, muitas vezes, você pega uma tese de 200 páginas e você quer localizar onde no texto o autor fala ‘disso’. Por exemplo: isso aqui parece que está nos métodos (e não está), aí, vamos para a avaliação de resultados (também não)...

Com certeza é útil, principalmente para quem não tem muita experiência em manusear trabalhos acadêmicos. É claro que nós estamos pensando aqui no caso de um pesquisador, mas, para muitas pessoas, por exemplo, da área empresarial (alguém que está buscando aquilo como subsídio para algum projeto que ele está desenvolvendo, ou alguma coisa assim) esse recurso auxilia na localização, pelo menos, das ideias-chave. Então, facilita a vida das pessoas.”

**Comentário 4 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“Qualquer ação para orientar o leitor, induzir o leitor a certas conclusões, a ler determinada parte em detrimento de outra é uma faca de dois gumes. Porque isso pode induzir os leitores mais apressados. Eles vão só ‘bater o olho’ naquela parte e deixar de fazer uma análise mais completa.

Em princípio, eu não vejo dificuldade, e eu acho até interessante, mas, isso pode viciar, pode apresentar uma certa tendenciosidade. Da mesma forma que ocorreu com a introdução do fator de impacto para avaliação de publicações e autores, eles viciaram, desviaram o processo... O pessoal usa uma série de artifícios.”

**Comentário 5 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“Quando o resumo é bem feito, você sabe se o artigo te interessa ou não e, mais além disso, a gente usa muito aqui com os alunos. Às vezes você está procurando uma referência e você trabalha só com o resumo, você lê e esse aqui é o trabalho de tal

aluno. Você já separa. Para mim é muito mais fácil, pode ser até questão de costume. No resumo estruturado, se a proposta for colocar... se o resumo for bem feito, talvez seja até melhor do que o tradicional, mas, você teria que se acostumar com esse formato. O resumo tradicional me dá o que o autor fez, como ele fez e que resultado ele obteve... isso é exatamente o que você precisa, e não precisa enumerar parágrafos nem nada, e, na verdade, a sequência será a do texto do autor. Ele tem que dar uma ou duas frases da introdução dele, depois na sequência, do corpo do trabalho e aqui ele vai sintetizar, depois, o resultado ele vai por aqui (mais uma ou duas frases), e a conclusão. É um bom resumo. Se interessou, você vai ler o trabalho dele, se o trabalho te interessou mais ainda, você vai nas referências do trabalho.”

**Comentário 6 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“Eu acho difícil, num trabalho mais complexo fica muito difícil você tentar fazer essa ‘referenciação’. Você imagina um trabalho como esse [o professor tomou um artigo em sua mesa], como é que você vai fazer isso? Em alguns tipos de artigos funciona, em outros, não.”

**Comentário 7 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“Vão ter situações em que uma sentença do resumo faça referência a vários pontos do artigo. Isso vai tumultuar a apresentação do resumo.”

**Comentário 8 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“O resumo deve representar o trabalho original, deve resumir de fato, não pode criar nada, ele é um item de representação. Com relação a essa questão de enumerar os parágrafos, pode ser interessante, pode ser interessante, desde que não crie confusão.

Se essa numeração, no resumo, for feita com cuidado para não ficar muito poluído o texto, é possível fazer essa associação. Isso dentro do resumo permite que o leitor encontre uma determinada informação resumida e decida: “deixa eu ver isso aqui mais em detalhe...”

Eu vejo com bons olhos o resumo estruturado, e acredito que possamos melhorar o resumo tradicional. É uma alternativa com aspectos interessantes (enumerar os parágrafos) e não me ocorre nenhuma outra sugestão para melhoria do resumo tradicional. É algo importante, buscaremos maneiras de melhorar os resumos.”

**Comentário 9 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“Com relação a essa enumeração, eu acho despropositada. Por exemplo, dependendo da área do artigo (se for em medicina, direito...) talvez seja necessário até. Mas, em princípio, eu acho um despropósito. Se você pega um resumo tradicional bem feito, ele é preciso, ele obedece a um padrão, ele é suficiente.”

**Comentário 10 sobre a sugestão de uso de “Indicação numérica de parágrafo”, no resumo, como auxílio na avaliação do texto original**

“Me parece uma tentativa de avanço. Num contexto muito consolidado como o atual acho que vale a pena esforços nesse sentido. Deve-se tomar cuidado, no entanto, para que o resumo não fique muito sujo ou ininteligível.”



A opinião dos professores e pesquisadores reproduzida acima ilustra algumas posições e alguns hábitos consolidados no uso de canais formais em comunicação científica. Durante a leitura de um determinado artigo os professores procuram avaliar o modo como o colega insere sua contribuição na área. Os periódicos científicos não são divididos em seções temáticas, agrupando os artigos que abordam um mesmo assunto (como fazem jornais e revistas não-científicas). Aspectos como esse fazem com que o valor do título, do resumo e das palavras-chave – no processo de avaliação do artigo – se torne mais objetivo. A atitude dos professores mostra muito cuidado com esses itens de identificação, particularmente em relação ao resumo, exatamente pelo valor discriminativo que eles possuem.

Com relação à enunciação da “originalidade e valor” do artigo junto ao resumo, os professores entrevistados, de forma geral, entendem que a questão de descrever o aspecto original de um artigo publicado é sempre delicada, mas, mesmo que não venha enunciada separadamente deveria constar implícita na própria redação escolhida para o resumo – já que se trata de um aspecto fundamental discutido no texto do documento original.

A indicação expressa de originalidade e valor responde a uma queixa levantada por Weinberg (1988) quando a autora afirma que a indexação frustra o pesquisador porque lida apenas de forma genérica com aquilo de que o documento trata e não lida, de modo algum, com o que esse documento proporciona de ‘novidade’ a respeito do assunto.

Iniciativas nesse sentido, realmente, são importantes porque uma indicação de originalidade e valor permite uma avaliação precisa da relevância do documento (veja HARTER, 1992). Em um ambiente em que publicações se proliferam freneticamente, quem acrescenta uma nova publicação ao corpo de conhecimento acumulado deveria deixar claro porque o fez.

Os aspectos inovadores e singulares de uma pesquisa, à primeira vista, deveriam saltar aos olhos do senso comum por sua própria natureza inusitada. A questão torna-se delicada porque, em um julgamento de valor assim, procuramos ponderar e distinguir *em que medida* um determinado aspecto inovador se afasta da prática já estabelecida.

É nessa atribuição de medida que muitos profissionais divergem, alguns avaliando a inovação positivamente e outros de uma forma mais restritiva.

A sugestão de indicação numérica do(s) parágrafo(s) onde a informação do resumo aparece discutida no original, no entanto, pareceu mais problemática. Foram mencionadas questões como: excesso de informação; correções que devem ser feitas na própria elaboração do resumo (antes da implementação de inovações); e, eventuais tendenciosidades no processo.

Essa sugestão foi apresentada com o intuito aumentar a objetividade e estreitar a relação entre um artigo e o seu sucedâneo, entretanto, a questão não se apresenta de maneira tão simples assim.

É difícil argumentar contra a observação de que, primeiro, correções devem ser feitas no próprio resumo antes da implementação de inovações. O resumo, enquanto texto, não pode desagregar o pensamento original que é coesivo. No processo de comunicação a representação proporcionada por um texto coerente é simplesmente virtuosa. Se o resumo falha em sua textualidade parece que somente uma inovação corrigirá essa falha: a redação de um novo texto que seja coeso e coerente.

O uso de indicação numérica de parágrafo(s) nos resumos levanta também questões ligadas à super-simplificação do processo comunicativo envolvido, podendo levar determinados usuários a inferir conclusões e atribuir significados de alcance reduzido, sem levar em devida consideração o conteúdo detalhado do texto original.

De outra forma, a indicação de parágrafo aproxima efetivamente os dois textos em um nível analítico mais avançado, tornando o exame do original de alguma forma mais estratégico. Essa sugestão, também, pode imprimir à noção de similaridade entre resumo e original alguma materialidade e objetividade.

## Apêndice B

### Corpus analisado nesta pesquisa

JDOC  
66,1

140

Received 12 August 2008  
Revised 5 June 2009  
Accepted 19 June 2009

#### SPECULATIONS IN DOCUMENTATION Information Literacy 2.0: hype or discourse refinement?

Sonja Špiranec and Mihaela Banek Zorica

*Department of Information Sciences, Faculty of Humanities and Social Sciences,  
University of Zagreb, Zagreb, Croatia*

##### Abstract

**Purpose** – The purpose of this paper is to introduce the term Information Literacy 2.0 as a subset of information literacy, to describe its development and significance and give an outline of its underlying theoretical assumptions.

**Design/methodology/approach** – The paper first examines the grounds for the possible re-conceptualizations in the field of information literacy and identifies the current developments in the information universe as the principal drive for perception shifts. Based on a literature review and a descriptive analysis of contrasting features of library user education, information literacy and Information literacy 2.0, the paper highlights the main foci of paradigm shifts.

**Findings** – The paper found that the new meaning and understanding of the central conceptions in information literacy are shifting the focus of classical information literacy towards Information Literacy 2.0. Many of the aspects of current information literacy practices originate from a print-based culture, which is incongruent with the transient and hybrid nature of digital environments. These radically changing environments are causing the appearance of anomalies in the information literacy paradigm, which could effectively be resolved through the introduction of a sub-concept of information literacy.

**Practical implications** – The article specifies the possibilities for putting theoretical conceptualizations of Information literacy 2.0 into practice by determining the range of shifts in information literacy activities and identifying how new practices differ from the earlier approaches.

**Originality/value** – The study attempts to advance the research field of information literacy by proposing a new outlook on information literacy through the integration of its underlying theoretical conceptions and practical applications.

**Keywords** Information, Literacy, Learning, Information facilities

**Paper type** Conceptual paper

RESUMO DISSERTATIVO

#### Grounded theory in practice: issues and discussion for new qualitative researchers

Jin Tan

*formerly Department of Information Studies, University of Sheffield, Sheffield,  
UK and now Sir Herbert Duthie Library, Cardiff University, Cardiff, UK*

Grounded theory  
in practice

93

Received 2 December 2008  
Revised 15 April 2009  
Accepted 17 April 2009

##### Abstract

**Purpose** – The purpose of this paper is to clarify four common confusions in grounded theory (GT) use by demonstrating an example of the author's doctoral study into conceptualising higher education students' learning and knowledge sharing by exploring blogging phenomena. It seeks not only to present the author's personal experience and views, but also to shed light on the causes of the confusion for novice qualitative researchers.

**Design/methodology/approach** – A review of literature on GT and the author's practical experience of undertaking an empirical study into the blogging phenomenon in LIS discipline form the approach to addressing the issue.

**Findings** – The paper summarises four common issues that hinder inexperienced qualitative researchers when they undertake GT research: using GT as a methodology or method; how to use literature review; how to use coding strategy; and what is the generated theory.

**Practical implications** – The paper provides practical suggestions of what matters when adopting GT approach. It needs more new researchers to further confirm the suggestions that the author stated in terms of a researcher's unique experience.

**Originality/value** – The paper attempts to bridge the gap of insufficient discussion in the literature, focuses on new researchers' GT adopting experiences and provides them with practical directions.

**Keywords** Literature, Codes

**Paper type** Research paper

RESUMO DISSERTATIVO

JDOC  
66,4

604

Received 24 July 2009  
Revised 9 November 2009  
Accepted 12 November 2009

## SPECULATIONS IN DOCUMENTATION

# The rhizome and the tree: changing metaphors for information organisation

Lyn Robinson and Mike Maguire

*Centre for Information Science, City University London, London, UK*

### Abstract

**Purpose** – The paper aims to review Deleuze and Guattari's concept of the rhizome as a model for information organisation.

**Design/methodology/approach** – This is a critical review of selected literature.

**Findings** – The rhizome concept is a promising model for understanding hyperlinked information services. It may be of practical value, particularly if it can be integrated with more traditional forms of information organisation. More research, conceptual and practical, is needed before this can be achieved.

**Research limitations/implications** – The literature review is not comprehensive, and the conclusions are open-ended.

**Originality/value** – This is the only paper to review the rhizome concept in this way.

**Keywords** Classification, Worldwide web, Philosophical concepts

**Paper type** Conceptual paper

### Introduction

At the heart of the information sciences lies the issue of classification (see, for example,

RESUMO DISSERTATIVO

## Classification in a social world: bias and trust

Jens-Erik Mai

*Faculty of Information, University of Toronto, Toronto, Canada*

Classification in  
a social world

627

### Abstract

**Purpose** – The purpose of this paper is to establish pluralism as the basis for bibliographic classification theory and practice and examine the possibility of establishing trustworthy classifications.

**Design/methodology/approach** – The paper examines several key notions in classification and extends previous frameworks by combining an explanation-based approach to classification with the concepts of cognitive authority and trust.

**Findings** – The paper presents an understanding of classification that allows designers and editors to establish trust through the principle of transparency. It demonstrates that modern classification theory and practice are tied to users' activities and domains of knowledge and that trustworthy classification systems are in close dialogue with users to handle bias responsibly and establish trust.

**Originality/value** – The paper establishes a foundation for exploring trust and authority for classification systems.

**Keywords** Classification, Philosophical concepts, Epistemology, Trust, Authority

**Paper type** Conceptual paper

Received 5 November 2009  
Revised 14 December 2009  
Accepted 15 December 2009

### Introduction

RESUMO DISSERTATIVO

# Document, text and medium: concepts, theories and disciplines

Niels Windfeld Lund  
*University of Tromsø, Tromsø, Norway*

734

Received 29 January 2009  
Revised 7 February 2010  
Accepted 11 February 2010

## Abstract

**Purpose** – The purpose of this paper is first to provide a critical conceptual discussion of different use of the notion of text, especially in the case of expressions including words as well as images, second to consider the notion of document as an alternative to the notion of text, and finally to lay out a theoretical ground for a broad discipline of documentation studies.

**Design/methodology/approach** – The approach took the form of a conceptual analysis of a number of works in media and literary studies.

**Findings** – There were found a number of cases of contradictory use of the notion of text within the same work, talking about text in a broad overall sense covering all media as well as text as a distinctive concept separating words from images, while it was found through a conceptual history of the notion of document that the latter notion not only covers a written paper, but multiple media.

**Research limitations/implications** – In future research, one should consider the use of document as the concept for the expressions as a whole and dedicate the notion of text solely for the verbal part of the expressions and make more empirical analysis within this conceptual framework to see if it makes a difference in practice to change the overall concept from text to document.

**Practical implications** – Having a broad concept of document, it would be possible to be more flexible regarding the choice of proper media for documentation.

**Originality/value** – By making a critique of the notion of text and suggesting a broader concept of document as well as documentation, the paper provides a ground for reconsidering the classical disciplinary structure, divided into humanities as well as the information sciences and social sciences.

**Keywords** Text retrieval, Document handling, Multimedia

RESUMO DISSERTATIVO

# The *h*-index: a broad review of a new bibliometric indicator

Michael Norris and Charles Oppenheim  
*Department of Information Science,  
Loughborough University, Loughborough, UK*

The *h*-index

681

## Abstract

**Purpose** – This review aims to show, broadly, how the *h*-index has become a subject of widespread debate, how it has spawned many variants and diverse applications since first introduced in 2005 and some of the issues in its use.

**Design/methodology/approach** – The review drew on a range of material published in 1990 or so sources published since 2005. From these sources, a number of themes were identified and discussed ranging from the *h*-index's advantages to which citation database might be selected for its calculation.

**Findings** – The analysis shows how the *h*-index has quickly established itself as a major subject of interest in the field of bibliometrics. Study of the index ranges from its mathematical underpinning to a range of variants perceived to address the indexes' shortcomings. The review illustrates how widely the index has been applied but also how care must be taken in its application.

**Originality/value** – The use of bibliometric indicators to measure research performance continues, with the *h*-index as its latest addition. The use of the *h*-index, its variants and many applications to which it has been put are still at the exploratory stage. The review shows the breadth and diversity of this research and the need to verify the veracity of the *h*-index by more studies.

**Keywords** Information studies, Indexing, Information research

**Paper type** Literature review

Received 19 June 2009  
Revised 5 January 2010  
Accepted 6 January 2010

RESUMO DISSERTATIVO

# Framing information literacy as information practice: site ontology and practice theory

Annemaree Lloyd

*School of Information Studies, Charles Sturt University, Wagga Wagga, Australia*

Framing  
information  
literacy

245

Received 8 April 2009  
Revised 12 July 2009  
Accepted 19 July 2009

## Abstract

**Purpose** – Information literacy is a rich and complex social information practice that is constructed according to specific practical understandings, rules and teleoaffective features which characterise a social site or setting. This paper aims to explore the philosophical and theoretical perspective of practice theory, in particular, the ontological work of Schatzki. These perspectives are to be used to frame an understanding of the features of information literacy as sociocultural practice.

**Design/methodology/approach** – A theoretical perspective is introduced to examine the concept of information literacy practice by framing this analysis through a site ontology developed by Schatzki. Sociocultural and practice theory are employed in this exploration of information literacy as sociocultural practice and provide a framework for architecture of information literacy practice.

**Findings** – Information literacy can be understood as a critical information practice which is organised and arranged through the site of the social, rather than as a reified and decontextualised set of skills.

**Research limitations/implications** – Framing information literacy research through site ontology and the use of a practice perspective has implications for further research into information literacy and for the development of pedagogic practices related to information literacy instruction

**Originality/value** – The paper offers an alternate way of framing information literacy by introducing the concepts related to practice theory.

**Keywords** Information literacy, Information management, Philosophical concepts

RESUMO DISSERTATIVO

JDOC  
66,1

28

Received 11 March 2009  
Revised 8 June 2009  
Accepted 29 June 2009

# The information needs of UK historic houses: mapping the ground

Alan Brine

*Library Services, De Montfort University, Leicester, UK, and*

John Feather

*Department of Information Science, Loughborough University, Loughborough, UK*

## Abstract

**Purpose** – The purpose of this paper is to concentrate on establishments that are known as the "built heritage", often described as "historic houses". Many operate in a similar vein to small businesses and often have a more diverse range of needs. In the UK historic houses form part of the tourism and leisure market. Heritage encompasses a wide variety of establishments including historic houses, historic gardens, heritage centres, town centres, countryside and museums.

**Design/methodology/approach** – A model was developed for the study to show the information needs of historic houses and the information-seeking behaviour of those managing historic houses. Data have been collected both via questionnaires to and interviews of those managing properties. The collection of data focused on diversity of activities, information provision, use of information, developments (including technology) and collaboration.

**Findings** – The conclusions discuss the management of information and the difficulties expressed for those managing historic houses as discovered during the research. Properties need to be able to make effective use of this to protect the heritage for future generations. Building on the empirical work recommendations are made on policy making, education, audits, co-operation and technology to improve the provision and management of information within the sector to support these establishments.

**Originality/value** – The research represents the first study to investigate the existing situation of historic houses and their information needs in the UK, to try and provide an overview of the sector and information provision and how that might be improved.

**Keywords** Heritage, Information management

**Paper type** Research paper

RESUMO DISSERTATIVO

# Turning weakness into strength: strategies for future LIS

Turning  
weakness into  
strength

Jan Nolin

*Swedish School of Library and Information Science, University of Borås,  
Borås, Sweden, and*

Fredrik Åström

*Lund University Libraries, Lund, Sweden and University of Technology Sydney,  
Sydney, Australia*

7

Received 20 February 2009  
Revised 9 June 2009  
Accepted 29 June 2009

## Abstract

**Purpose** – LIS has been described as a fragmented field in crisis, with an increased competition from other fields; and lacking in development of theories. The purpose of this paper is to articulate a strategy in which the perceived weakness can be seen as a source of strength.

**Design/methodology/approach** – The text builds mostly on reflections on meta-theoretical and science-organisation literature. Ten distinct problems for the research field are identified and discussed in order to provide a viable strategy for the future.

**Findings** – While it is common to suggest a convergent movement toward the idealised characteristics of the strong research discipline as a recipe against fragmentation, a strong convergent movement is suggested that feeds off the fragmented character of the field. What is commonly perceived as a weakness, the multidimensional character of the field, can be translated into a strategic resource.

**Originality/value** – The paper provides a fresh perspective on the strategic situation of LIS.

**Keywords** Research work, Information science, Uncertainty management

**Paper type** Viewpoint

RESUMO EXPOSITIVO

# A preservation framework for Chinese ancient books

A preservation  
framework for  
Chinese books

Mingjie Li

*School of Information Management, Wuhan University, Wuhan, China, and*

Jinfang Niu

*School of Information Sciences, University of Tennessee at Knoxville,  
Knoxville, Tennessee, USA*

259

Received 4 December 2008  
Revised 1 June 2009  
Accepted 22 July 2009

## Abstract

**Purpose** – This paper aims to provide a theoretical guide for preserving ancient books in China.

**Design/methodology/approach** – The paper reviews the history of the damage and preservation of ancient books in China, and analyzes the value attributes of ancient books: archaeological value, historical value and artistic value.

**Findings** – The paper proposes a preservation framework for Chinese ancient books. This framework is composed of three layers. The foundation layer is to preserve the physical entity of ancient books so that the archaeological values are preserved. The middle layer is to preserve the intellectual content of ancient books so that the values for historical research are preserved. The top layer is mainly about preserving the productions process of the artistic format of ancient books, so that not only the static artistic formats are preserved, the techniques and procedures to produce the artistic format are preserved as well.

**Originality/value** – The paper presents a framework that connects the value attributes of ancient books and the strategies to preserve those values, systematizes them and presents them as a whole. The framework can be used to justify government policies and help identify pitfalls in the preservation strategies for ancient books.

**Keywords** Books, China, Historical research, Collections management

**Paper type** Conceptual paper

RESUMO EXPOSITIVO

## Developing scales for information-seeking behaviour

Caroline F. Timmers

*Institute of Marketing & International Management,  
Saxion University of Applied Sciences, Enschede, The Netherlands, and*

Cees A.W. Glas

*Department of Research Methodology, Measurement and Data Analysis,  
Faculty of Behavioral Sciences, University of Twente, Enschede,  
The Netherlands*

### Abstract

**Purpose** – The main purpose of this paper is to describe the development of an instrument designed to measure information-seeking behaviour of undergraduate students during study assignments.

**Design/methodology/approach** – Literature research, internal consistency and reliability computed with Cronbach's Alpha ( $\alpha$ ), Factor Analyses with Varimax rotation, and item response theory form the approach to examining the subject.

**Findings** – Four scales were found within a 46-item survey on information-seeking behaviour: a ten-item scale for applying search strategies ( $\alpha = 0.68$ ), a 14-item scale for evaluating information ( $\alpha = 0.74$ ), a six-item scale for referring to information ( $\alpha = 0.81$ ) and a 12-item scale for regulation activities when seeking information ( $\alpha = 0.75$ ).

**Originality/value** – The four scales for information-seeking behaviour can be used to monitor and evaluate this behaviour of students in higher education.

**Keywords** Information, Literacy, Information research, Higher education, The Netherlands

**Paper type** Research paper

RESUMO EMPÍRICO

## Source preference criteria in the context of everyday projects Relevance judgments made by prospective home buyers

Reijo Savolainen

*Department of Information Studies and Interactive Media,  
University of Tampere, Tampere, Finland*

### Abstract

**Purpose** – The purpose of this paper is to elaborate how source preference criteria are defined in the context of everyday projects that require the seeking of problem-specific information. More specifically, to find out how information seekers explain their preference criteria by characterizing the perceived strengths and weaknesses of diverse sources.

**Design/methodology/approach** – The approach takes the form of qualitative content analysis of empirical data gathered by semi-structured interviews with 16 prospective home buyers in 2008. The source preference criteria were elicited by making use of the construct of information source horizon.

**Findings** – Networked sources were favoured most strongly, followed by printed media, human sources and organizational sources. Content of information was the primary source preference criterion. Availability of information was a fairly important criterion, while user characteristics, usability of information and situational factors were fairly marginal in this regard. In the definition of the preference criteria, more emphasis was placed on the perceived strengths than weaknesses of sources. Positive qualities such as "provides updated information" were referred to particularly while judging the relevance of the networked sources. Negative qualities like "outdated information" were primarily associated with printed media and organizational sources.

**Research limitations/implications** – The study is exploratory, drawing on a relatively small sample recruited through a web-based service. Thus, the findings cannot be generalized to prospective home buyers.

**Practical implications** – Prospective home buyers tend to favour web-based information sources and services. They should provide the customers with detailed information about the property, including photos.

**Originality/value** – The paper specifies the picture of user-defined relevance judgment in the context of everyday life information seeking.

**Keywords** Information retrieval, Information media

RESUMO EMPÍRICO

# Information behaviour of women: theoretical perspectives on gender

Information  
behaviour of  
women

113

Christine Urquhart and Alison Yeoman

*Department of Information Studies, Aberystwyth University, Aberystwyth, UK*

Received 29 August 2008  
Revised 9 April 2009  
Accepted 15 April 2009

## Abstract

**Purpose** – The purpose of this paper is to examine whether there is a need to consider gender or sex differences as variables in information behaviour research and, if so, how?

**Design/methodology/approach** – A metasynthesis approach is used. A preliminary framework to categorise information behaviour research on women is developed by integrating main themes from feminist research and information behaviour research. Within each category, studies are compared and contrasted, to identify similar and divergent themes. Themes are then compared across categories, to synthesise the main concepts.

**Findings** – The categorisation works for most studies, apart from a group of studies on health information use, communicating risk and decision making. The meta-synthesis indicates the importance of concepts such as situation (as mesh), intermediaries (as node with connections), and connecting behaviour. Gender-related or, rather gender-ascribed, constructs, such as concern for others, not gender alone are likely to be important variables in information behaviour.

**Research limitations/implications** – The meta-synthesis is a top-level synthesis, as the number of studies prohibited a more detailed approach. Further meta-synthesis of a few high quality research studies would help to confirm the findings.

**Practical implications** – The synthesis illuminates a different perspective on information behaviour: the network of information users rather than the individual information seeker.

**Originality/value** – The synthesis integrates some feminist research themes with information behaviour research, and the findings have implications for general information behaviour research.

RESUMO EMPÍRICO

# On the prevalence and scientific impact of duplicate publications in different scientific fields (1980-2007)

On the prevalence  
of duplicate  
publications

179

Vincent Larivière and Yves Gingras

*Observatoire des sciences et des technologies (OST),*

*Centre interuniversitaire de recherche sur la science et la technologie (CIRST),  
Université du Québec à Montréal, Montréal, Canada*

## Abstract

**Purpose** – The issue of duplicate publications has received a lot of attention in the medical literature, but much less in the information science community. This paper aims to analyze the prevalence and scientific impact of duplicate publications across all fields of research between 1980 and 2007.

**Design/methodology/approach** – The approach is a bibliometric analysis of duplicate papers based on their metadata. Duplicate papers are defined as papers published in two different journals having: the exact same title; the same first author; and the same number of cited references.

**Findings** – In all fields combined, the prevalence of duplicates is one out of 2,000 papers, but is higher in the natural and medical sciences than in the social sciences and humanities. A very high proportion (>85 percent) of these papers are published the same year or one year apart, which suggest that most duplicate papers were submitted simultaneously. Furthermore, duplicate papers are generally published in journals with impact factors below the average of their field and obtain lower citations.

**Originality/value** – The paper provides clear evidence that the prevalence of duplicate papers is low and, more importantly, that the scientific impact of such papers is below average.

**Keywords** Ethics, Medical sciences, Intellectual property, Publications, Information science

**Paper type** Research paper

RESUMO EMPÍRICO

# Exploring young children's web searching and technoliteracy

Young children's  
web searching

Amanda Spink, Susan Danby, Kerry Mallan and Carly Butler  
*Queensland University of Technology, Brisbane, Queensland, Australia*

191

Received 26 November 2008  
Revised 4 July 2009  
Accepted 6 July 2009

## Abstract

**Purpose** – This paper aims to report findings from an exploratory study investigating the web interactions and technoliteracy of children in the early childhood years. Previous research has studied aspects of older children's technoliteracy and web searching; however, few studies have analyzed web search data from children younger than six years of age.

**Design/methodology/approach** – The study explored the Google web searching and technoliteracy of young children who are enrolled in a "preparatory classroom" or kindergarten (the year before young children begin compulsory schooling in Queensland, Australia). Young children were video- and audio-taped while conducting Google web searches in the classroom. The data were qualitatively analysed to understand the young children's web search behaviour.

**Findings** – The findings show that young children engage in complex web searches, including keyword searching and browsing, query formulation and reformulation, relevance judgments, successive searches, information multitasking and collaborative behaviours. The study results provide significant initial insights into young children's web searching and technoliteracy.

**Practical implications** – The use of web search engines by young children is an important research area with implications for educators and web technologies developers.

**Originality/value** – This is the first study of young children's interaction with a web search engine.

**Keywords** Australia, Worldwide web, Information searches, Children (age groups), Information literacy, Behaviour

Paper type Research paper

RESUMO EMPÍRICO

# From digital libraries to digital preservation research: the importance of users and context

From libraries to  
preservation  
research

Gobinda Chowdhury  
*University of Technology, Sydney, Australia*

207

Received 1 January 2009  
Revised 10 July 2009  
Accepted 17 July 2009

## Abstract

**Purpose** – The purpose of this paper is to point out the commonalities of research in digital libraries and digital preservation with regard to the issues of users and context of information.

**Design/methodology/approach** – The paper's approach is a review of selected literature and reports of research projects focusing particularly on digital preservation research.

**Findings** – It is noted that just like the digital library community the digital preservation research community is also confronted with the challenges of capturing, storing and making use of the information related to users and context.

**Practical implications** – The paper points out some current research in digital preservation that aims to handle the users and context information for building future digital preservation systems. It highlights some major challenges in these areas.

**Originality/value** – The paper reports on the state of the art research in digital preservation.

**Keywords** Digital libraries, Information science, Collections management, Information management, User studies

Paper type Literature review

RESUMO EMPÍRICO

# The connection between purposive information seeking and information encountering

## A study of Icelanders' health and lifestyle information seeking

Ágústa Pálsdóttir

*Department of Library and Information Science, University of Iceland,  
Reykjavik, Iceland*

### Abstract

**Purpose** – This paper aims to explore health and lifestyle information seeking behaviour by examining the connection between purposive information seeking and information encountering.

**Design/methodology/approach** – Data on purposive seeking and information encountering, gathered from postal surveys in 2002 and 2007, were compared. Random samples of 1,000 Icelanders, aged 18 to 80 were used. The response rate was 51 per cent in 2002 and 47 per cent in 2007. Based on the purposive seeking in 22 sources, *k*-means cluster analysis was used to draw four clusters of participants: passive, moderately passive, moderately active and active.

**Findings** – The results from 2007 and 2002 revealed the same kind of information seeking. The findings indicate that information encountering is an integral feature of information seeking behaviour. Information is encountered more often than sought on purpose by all clusters. Clusters that were active in purposive information seeking were also active in information encountering and those who were passive in either of the two styles of information seeking were also passive in the other.

**Research limitations/implications** – The response rates are considered satisfactory in postal surveys. Nevertheless, when missing data in the cluster analysis are also considered it raises a question about the validity of the findings. The findings of the studies, however, are strengthened by the fact that respondents reflect the population fairly well.

**Practical implications** – Improved knowledge of information seeking and how different groups within society can be reached more effectively is important for health promotion and public health practice.

**Originality/value** – The paper uses quantitative methods to examine the connection between purposive information seeking and information encountering.

**Keywords** Personal health, Lifestyles, Iceland, Information searches, Individual behaviour,

RESUMO EMPÍRICO

# The effect of dyslexia on information retrieval

## A pilot study

A. MacFarlane

*Centre for Interactive Systems Research, City University London, London, UK*

A. Al-Wabil

*Centre for HCI Design, City University London, London, UK*

C.R. Marshall

*Department of Language and Communication Science, City University London,  
London, UK*

A. Albrair and S.A. Jones

*Centre for Interactive Systems Research, City University London, London, UK, and*

P. Zaphiris

*Centre for HCI Design, City University London, London, UK*

The effects of  
dyslexia on IR

307

Received 15 December 2008  
Revised 7 August 2009  
Accepted 11 August 2009

### Abstract

**Purpose** – The purpose of this paper is to resolve a gap in the knowledge of how people with dyslexia interact with information retrieval (IR) systems, specifically an understanding of their information-searching behaviour.

**Design/methodology/approach** – The dyslexia cognitive profile is used to design a logging system, recording the difference between two sets of participants: dyslexic and control users. A standard Okapi interface is used – together with two standard TREC topics – in order to record the information searching behaviour of these users.

**Findings** – Using the log data, the differences in information-searching behaviour of control and dyslexic users, i.e. in the way the two groups interact with Okapi, are established and it also established that qualitative information collected (such as experience etc.) may not be able to account for these differences. Evidence from query variables was unable to distinguish between groups, but differences on topic for the same variables were recorded. Users who view more documents tended to judge more documents as being relevant, in terms of either the user group or topic. Session data indicated that there may be an important difference between the number of iterations used in a search between the user groups, as there may be little effect from the topic on this variable.

**Originality/value** – This is the first study of the effect of dyslexia on information search behaviour, and it provides some evidence to take the field forward.

**Keywords** Information retrieval, Dyslexia, Behaviour, User interfaces

**Paper type** Research paper



Journal of Documentation  
Vol. 66, No. 3, 2010

The authors are grateful to Helen Petrie for her input on the initial part of this research, in

RESUMO EMPÍRICO

# Creativity and craft: the information-seeking behavior of theatre artists

Creativity and  
craft

Ann Medaille

*Mathewson-IGT Knowledge Center, University of Nevada, Reno, Nevada, USA*

327

Received 26 January 2009  
Revised 10 August 2009  
Accepted 11 August 2009

## Abstract

**Purpose** – The aim of this study is to explore the information needs and behaviors of practicing theatre artists. Psychological research into creativity provides a framework for understanding both theatre artists' information-seeking behavior and the role of information seeking and gathering in the creative process.

**Design/methodology/approach** – The exploratory study presents findings from an online questionnaire of 73 practicing theatre artists and qualitative data gathered from eight interviews with theatre professionals.

**Findings** – The study reveals that theatre artists seek information for six primary purposes: understanding a work's historical, cultural, and critical background; finding sources of inspiration; learning about contemporary or historical theatre productions, artists, and events; learning technical or process information; finding performance materials; and furthering career goals. Theatre artists view the information search process as being essential to their creative activities, and their first-hand accounts of their artistic experiences illuminate the critical role that information seeking and gathering play in the creative process.

**Research limitations/implications** – Some theatre professions, such as lighting or sound design, were represented in the questionnaire but were not represented in the interviews.

**Practical implications** – The study has practical implications for the delivery of library and internet theatre art collections and information services.

**Originality/value** – Few studies have examined the information-seeking behavior of practicing theatre artists. The paper demonstrates that studies of artists can be used to understand the role of information seeking and gathering in the creative processes of people working in various subject domains.

**Keywords** Arts, Information research, Theatre, Behaviour

Paper type Research paper

RESUMO EMPÍRICO

JDOC  
66,3

# Information about the synthesis of life forms: a document-oriented approach

348

Received 4 May 2009  
Revised 11 August 2009  
Accepted 12 August 2009

Inna Kouper

*School of Library and Information Science, Indiana University, Bloomington,  
Indiana, USA*

## Abstract

**Purpose** – This study aims to examine the forms of information about the synthesis of life forms in the public sphere.

**Design/methodology/approach** – A document-oriented approach was used and a wide range of documents that discuss a particular technoscientific issue was sampled. The analysis of documents involved a combination of discourse and content analysis.

**Findings** – The study demonstrates that there is a significant growth of the diversity of document types over time. Overall, 24 document types and 21 publication formats were identified. Web-based formats, such as blogs and news and information web sites, play a prominent role in the dissemination of information about the synthesis of life forms.

**Research limitations/implications** – The variety of document types identified here expands current understanding of the public documentary landscape and shows that the analysis of technoscientific debates and controversies can no longer be limited to traditional mass media documents such as news, feature articles, and editorials. However, a larger sample that includes more documents as well as non-textual objects, such as images or even lab specimens, would expand the scope of this taxonomy and make conclusions more definitive. Further research into the specific digital types of documents identified in the study and their impact on the communication of scientific information to the public is needed.

**Practical implications** – Surveying and understanding the kinds of documents that circulate information about emerging technoscientific issues can help to provide better services for a variety of information users and develop better tools for access and dissemination of such information.

**Originality/value** – The study demonstrates that a document-oriented approach can provide valuable insight into the circulation of information about science in the public sphere. It also offers an elaborate taxonomy of documents that can be used in further research as well as in information and science literacy instruction.

**Keywords** Information science, Communication technologies,

RESUMO EMPÍRICO

## Orally-based information

Deborah Turner

*Information School, University of Washington, Seattle, Washington, USA*

370

Received 15 March 2009  
Revised 18 August 2009  
Accepted 20 August 2009

### Abstract

**Purpose** – The purpose of this paper is to explore a new research area: orally-based information.

**Design/methodology/approach** – The study utilizes a social constructionist approach. The social constructionist meta-theory, which holds that contributions to knowledge can be made orally, frames it.

**Findings** – The paper explicates how orality, or word-of-mouth transactions, conveys information; describes approaches for investigating orally-based information; and articulates the need for future information behavior investigations that focus on orality.

**Research limitations/implications** – The research exploration focuses on face-to-face oral data. It calls for increased attention to orally-based information, and offers tentative suggestions for accomplishing this goal.

**Practical implications** – The results provide insight that assist in understanding how orally-based information intersects with information behavior, knowledge management, information policy, cultural heritage, and professional development that involves orality.

**Originality/value** – The paper builds a theoretical foundation for increased understanding of the meaning and functions of orally-based information.

**Keywords** Interpersonal communications, Information exchange, Face-to-face communications, Information science

**Paper type** Conceptual paper

### RESUMO EMPÍRICO

## Scholarly output: print and digital – in teaching and research

Sally Maynard and Ann O'Brien

*Department of Information Science, Loughborough University,  
Loughborough, UK*

384

Received 19 November 2008  
Revised 20 August 2009  
Accepted 20 August 2009

### Abstract

**Purpose** – The purpose of this paper is to report the outcomes of a Joint Information Systems Committee (JISC)-sponsored study to determine the current state and trends in different forms of scholarly output used in teaching and research; and the nature and extent of problems associated with their use.

**Design/methodology/approach** – A total of 60 UK HE institutions were chosen at random and a selection of departments within these was contacted. An online questionnaire was distributed to the selected departments; resulting in responses from 304 academics across a broad range of subjects and institution types.

**Findings** – The study showed that printed output was still the preferred option in both teaching and research, although electronic journals now have a well-established presence. Web-based material is increasingly provided in teaching and used in research but this includes primarily traditional tools such as reading lists and links to scholarly resources. Some content creation was evident. Use of web 2.0 was not extensive, although respondents were making use of Institutional Learning Environments. Academics were aware of Intellectual Property Rights (IPR) issues but not always clear about their responsibilities in this area.

**Research limitations/implications** – The study revealed an essentially conservative approach to the developments in digital information. This may have been due to the sample size which was relatively small, and the age profile which clustered around the 45-65 years range. In the case of research the influence of the Research Assessment Exercise (RAE) was clear.

**Originality/value** – No equivalent study has been reported on the transition between traditional and new forms of scholarly output used in teaching and research. In this fast developing area this research provides a benchmark for future studies.

**Keywords** Print media, Electronic journals, Teaching, Research, Intellectual property, Computer applications

**Paper type** Research paper

### Introduction

### RESUMO EMPÍRICO

## Diversity in the e-journal use and information-seeking behaviour of UK researchers

E-journal use of UK researchers

409

David Nicholas, Ian Rowlands and Paul Huntington  
*CIBER, Department of Information Studies, University College London, London, UK*

Hamid R. Jamali  
*Department of Library and Information Studies, Faculty of Psychology and Education, Tarbiat Moallem University, Tehran, Iran, and*

Patricia Hernández Salazar  
*University Centre for Library Science Research, National Autonomous University of México, Mexico City, México*

Received 2 April 2009  
Revised 25 August 2009  
Accepted 26 August 2009

### Abstract

**Purpose** – The purpose of this paper is to present some of the results of the project ‘Evaluating the usage and impact of e-journals in the UK’. The particular research reported here evaluated the use of the ScienceDirect journals database with regard to Life Sciences, Economics, Chemistry, Earth & Environmental Sciences and Physics by ten major UK research institutions. The aim of the study is to investigate researchers’ digital behaviour, and to ascertain whether their use and behaviours varied by subjects and disciplines, or in relation to the institutions in which they worked.

**Design/methodology/approach** – Raw logs for ScienceDirect were obtained for the period January to April 2007, were subject to deep log techniques and analysed using the Software Package for Social Sciences (SPSS).

**Findings** – Typically, 5 per cent of the ScienceDirect journals viewed accounted for a third to half of all use. A high proportion of researchers entered the ScienceDirect site via a third-party site, and this was especially so in the case of the Life Sciences and in the highest-ranked research institutions. There were significant institutional and subject differences in information-seeking behaviour. In the most research-intensive institutions, per capita journal use was highest and their users spent much less time on each visit. There were significant differences of the order of 100–300 per cent in the age of material viewed between subjects and institutions. Just four months after ScienceDirect content was opened to Google indexing, a third of traffic to the site’s Physics journals came via that route.

**Originality/value** – The research is one of the very few studies to investigate subject and institutional differences with regard to the information seeking and use of UK researchers, something UK academic librarians should particularly welcome.

**Keywords** Information research, Consumer behaviour, Electronic journals, United Kingdom

**Paper type** Research paper

### Introduction

Numerous studies testify to the fact that increased and enhanced online access to the



Journal of Documentation  
Vol. 66, No. 3, 2010

RESUMO EMPÍRICO

JDOC  
66,3

## Analysis of errors in the automatic translation of questions for translingual QA systems

434

Received 12 February 2009  
Revised 2 September 2009  
Accepted 3 September 2009

María-Dolores Olvera-Lobo  
*CSIC, Unidad Asociada Grupo SCImago, Madrid, España, and*

Lola García-Santiago  
*Department of Library and Information Science, University of Granada, Granada, Spain*

### Abstract

**Purpose** – This study aims to focus on the evaluation of systems for the automatic translation of questions destined to translingual question-answer (QA) systems. The efficacy of online translators when performing as tools in QA systems is analysed using a collection of documents in the Spanish language.

**Design/methodology/approach** – Automatic translation is evaluated in terms of the functionality of actual translations produced by three online translators (Google Translator, Prompt Translator, and Worldlingo) by means of objective and subjective evaluation measures, and the typology of errors produced was identified. For this purpose, a comparative study of the quality of the translation of factual questions of the CLEF collection of queries was carried out, from German and French to Spanish.

**Findings** – It was observed that the rates of error for the three systems evaluated here are greater in the translations pertaining to the language pair German-Spanish. Prompt was identified as the most reliable translator of the three (on average) for the two linguistic combinations evaluated. However, for the Spanish-German pair, a good assessment of the Google online translator was obtained as well. Most errors (46.38 percent) tended to be of a lexical nature, followed by those due to a poor translation of the interrogative particle of the query (31.16 percent).

**Originality/value** – The evaluation methodology applied focuses above all on the finality of the translation. That is, does the resulting question serve as effective input into a translingual QA system? Thus, instead of searching for ‘perfection’, the functionality of the question and its capacity to lead one to an adequate response are appraised. The results obtained contribute to the development of improved translingual QA systems.

**Keywords** Translation services, Computer applications, Knowledge management, Languages, Error analysis, Quality improvement

RESUMO EMPÍRICO

# Imagining the internet: learning and access to information in Singapore's public libraries

Adrian Kay Heng Heok and Brendan Luyt  
*Wee Kim Wee School of Communication and Information,  
Nanyang Technological University, Singapore*

Imagining the  
internet

475

Received 7 June 2009  
Revised 17 September 2009  
Accepted 21 September  
2009

## Abstract

**Purpose** – This study aims to investigate the discourses surrounding the provision of internet access in Singapore through the public library system and to consider what the government, librarians and members of the public consider are the legitimate uses of the internet in Singapore's public libraries, how these compare with what the librarians see as their role in facilitating access and to what extent the internet as an educational tool features in public libraries according to users.

**Design/methodology/approach** – A discourse analysis of public documents and semi-structured interviews with ten senior librarians, managers, and 40 members of the public were conducted. Observation was also conducted of library users and of their usage at the computer terminals in the library.

**Findings** – Findings from the study throw up very different definitions of what constitutes learning through the internet within the context of a public library. This reinforces conclusions from previous studies that what is termed learning through the internet is variously constructed, and understood in multiple ways. This defies what is often presented of the internet as a technical quick fix of policy makers to help its population leap frog into the future.

**Research limitations/implications** – While the study provided a glimpse into the discourse surrounding the internet in one country, similar work should be carried in more Asian countries to provide a more comprehensive survey of the region and compare this with findings conducted in an Occidental setting.

**Originality/value** – The study provides an Asian perspective that complements findings on the various discourses that had surrounded internet technology in the USA and Europe. It also calls for a new spirit of awareness on the part of librarians to the views of library users with regard to internet access.

**Keywords** Public libraries, Singapore, Internet, Government, Librarians, Library users

**Paper type** Research paper

RESUMO EMPÍRICO

# Two kinds of evidence: how information systems form rhetorical arguments

Melanie Feinberg  
*Information School, University of Texas, Austin, Texas, USA*

Two kinds of  
evidence

491

Received 4 March 2009  
Revised 22 September 2009  
Accepted 23 September  
2009

## Abstract

**Purpose** – This paper aims to examine how systems for organizing information construct rhetorical arguments for a particular interpretation of their subject matter.

**Design/methodology/approach** – The paper synthesizes a conceptual framework from the field of rhetoric and uses that framework to analyze how existing organizational schemes present evidence in support of arguments regarding the material being organized.

**Findings** – Organizational schemes can present logical arguments as posed in rhetoric, using two forms of evidence for their claims: relationship evidence from the category structure and resource evidence from the ways that items are assigned to categories.

**Research limitations/implications** – This study does not attempt to identify all types of evidence that organizational schemes might use in argumentation. Further research may describe additional forms of evidence and argumentative structures.

**Practical implications** – When creating organizational schemes, designers might develop a strategy to facilitate persuasive argumentation. Moreover, because arguments may be either strengthened or undermined through the assignment of resources to categories, both indexing and collection development may be seen as contributing to the overall design of an organizational scheme.

**Originality/value** – While many researchers have asserted that organizational schemes form arguments, and while various studies have described what information systems might be said to communicate, this study focuses on how such communication may take place more or less effectively. This analysis foregrounds the potential for organizational schemes to be systematically and purposefully designed as rhetorical communication, to express particular ideas.

**Keywords** Knowledge organizations, Classification, Information systems, Design

**Paper type** Research paper

RESUMO EMPÍRICO

# An experimental comparison of a genetic algorithm and a hill-climber for term selection

An experimental comparison

513

A. MacFarlane

*Centre for Interactive Systems Research, Department of Information Science,  
City University, London, UK*

A. Secker and P. May

*Computing Laboratory, University of Kent, Canterbury, UK, and*

J. Timmis

*Departments of Electronics and Computer Science, University of York, York, UK*

## Abstract

**Purpose** – The term selection problem for selecting query terms in information filtering and routing has been investigated using hill-climbers of various kinds, largely through the Okapi experiments in the TREC series of conferences. Although these are simple deterministic approaches, which examine the effect of changing the weight of one term at a time, they have been shown to improve the retrieval effectiveness of filtering queries in these TREC experiments. Hill-climbers are, however, likely to get trapped in local optima, and the use of more sophisticated local search techniques for this problem that attempt to break out of these optima are worth investigating. To this end, this paper aims to apply a genetic algorithm (GA) to the same problem.

**Design/methodology/approach** – A standard TREC test collection is used from the TREC-8 filtering track, recording mean average precision and recall measures to allow comparison between the hill-climber and GAs. It also varies elements of the GA, such as probability of a word being included, probability of mutation and population size in order to measure the effect of these variables. Different strategies such as elitist and non-elitist methods are used, as well as roulette wheel and rank selection GAs.

**Findings** – The results of tests suggest that both techniques are, on average, better than the baseline, but, the implemented GA does not match the overall performance of a hill-climber. The Rank selection algorithm does better on average than the Roulette Wheel algorithm. There is no evidence in this study that varying word inclusion probability, mutation probability or Elitist method make much difference to the overall results. Small population sizes do not appear to be as effective as larger population sizes.

**Research limitations/implications** – The evidence provided here would suggest that being stuck in a local optima for the term selection optimization problem does not appear to be detrimental to the overall success of the hill-climber. The evidence from term rank order would appear to provide extra useful evidence, which hill-climbers can use efficiently, and effectively, to narrow the search space.

**Originality/value** – The paper represents the first attempt to compare hill-climbers with GAs on a problem of this type.

**Keywords** Information searches, Programming and algorithm theory, Retrieval performance evaluation, Indexing

**Paper type** Research paper



Journal of Documentation

RESUMO EMPÍRICO

JDOC  
66,4

532

Received 30 September 2008  
Revised 5 May 2009  
16 November 2009  
Accepted 22 November 2009

# PERCIRS: a system to combine personalized and collaborative information retrieval

Hassan Naderi

*Department of Computer Engineering,  
Iran University of Science and Technology (IUST), Tehran, Iran, and*

Beatrice Rumpler

*Université de Lyon, CNRS, Insa-Lyon, Liris, France*

## Abstract

**Purpose** – This paper aims to discuss and test the claim that utilization of the personalization techniques can be valuable to improve the efficiency of collaborative information retrieval (CIR) systems.

**Design/methodology/approach** – A new personalized CIR system, called PERCIRS, is presented based on the user profile similarity calculation (UPSC) formulas. To this aim, the paper proposes several UPSC formulas as well as two techniques to evaluate them. As the proposed CIR system is personalized, it could not be evaluated by Cranfield, like evaluation techniques (e.g. TREC). Hence, this paper proposes a new user-centric mechanism, which enables PERCIRS to be evaluated. This mechanism is generic and can be used to evaluate any other personalized IR system.

**Findings** – The results show that among the proposed UPSC formulas in this paper, the (query-document)-graph based formula is the most effective. After integrating this formula into PERCIRS and comparing it with nine other IR systems, it is concluded that the results of the system are better than the other IR systems. In addition, the paper shows that the complexity of the system is less than the complexity of the other CIR systems.

**Research limitations/implications** – This system asks the users to explicitly rank the returned documents, while explicit ranking is still not widespread enough. However it believes that the users should actively participate in the IR process in order to aptly satisfy their needs to information.

**Originality/value** – The value of this paper lies in combining collaborative and personalized IR, as well as introducing a mechanism which enables the personalized IR system to be evaluated. The proposed evaluation mechanism is very valuable for developers of personalized IR systems. The paper also introduces some significant user profile similarity calculation formulas, and two techniques to evaluate them. These formulas can also be used to find the user's community in the social networks.

**Keywords** Information retrieval, Social networks, Customization, Information systems

RESUMO EMPÍRICO

# Perspectives of studies on document abstracting

## Towards an integrated view of models and theoretical approaches

Mónica Izquierdo Alonso

*Faculty of documentation, Alcalá University, Alcalá, Madrid, Spain, and*

Luis Miguel Moreno Fernández

*Department of Information and Documentation, University of Murcia, Murcia, Spain*

Studies on document abstracting

563

Received 29 June 2009  
Revised 11 November 2009  
Accepted 22 November 2009

### Abstract

**Purpose** – The aim of this paper is to systemize and improve the scientific status of studies on document abstracting. This is a diachronic, systematic study of document abstracting studies carried out from different perspectives and models (textual, psycholinguistic, social and communicative).

**Design/methodology/approach** – A review of the perspectives and analysis proposals which are of interest to the various theoreticians of abstracting is carried out using a variety of techniques and approaches (cognitive, linguistic, communicative-social, didactic, etc.), each with different levels of theoretical and methodological abstraction and degrees of application. The most significant contributions of each are reviewed and highlighted, along with their limitations.

**Findings** – It is found that the great challenge in abstracting is the systemization of models and conceptual apparatus, which open up this type of research to semiotic and socio-interactive perspectives. It is necessary to carry out suitable empirical research with operative designs and *ad hoc* measuring instruments which can measure the efficiency of the abstracting and the efficiency of a good abstract, while at the same time feeding back into the theoretical baggage of this type of study. Such research will have to explain and provide answers to all the elements and variables, which affect the realization and the reception of a quality abstract.

**Originality/value** – The paper provides a small map of the studies on document abstracting. This shows how the conceptual and methodological framework has extended at the same time as the Science of Documentation has been evolving. All the models analysed – the communicative and interactional approach – are integrated in a new systematic framework.

**Keywords** Abstracting, Document management, Information retrieval, Research

RESUMO EMPÍRICO

# Cognitive styles and search engine preferences

## Field dependence/independence vs holism/serialism

Natalie Clewley

*School of Information Systems, Computing, and Mathematics, Brunel University, Uxbridge, UK*

Sherry Y. Chen

*School of Information Systems, Computing, and Mathematics, Brunel University, Uxbridge, UK and Graduate Institute of Network Learning Technology, National Central University, Jhongli City, Taiwan, and*

Xiaohui Liu

*School of Information Systems, Computing, and Mathematics, Brunel University, Uxbridge, UK*

Cognitive styles and preferences

585

Received 8 February 2009  
Revised 4 December 2009  
Accepted 6 December 2009

### Abstract

**Purpose** – Cognitive style has been identified to be significantly influential in deciding users' preferences of search engines. In particular, Witkin's field dependence/independence has been widely studied in the area of web searching. It has been suggested that this cognitive style has conceptual links with the holism/serialism. This study aims to investigate the differences between the field dependence/independence and holism/serialism.

**Design/methodology/approach** – An empirical study was conducted with 120 students from a UK university. Riding's cognitive style analysis (CSA) and Ford's study preference questionnaire (SPQ) were used to identify the students' cognitive styles. A questionnaire was designed to identify users' preferences for the design of search engines. Data mining techniques were applied to analyse the data obtained from the empirical study.

**Findings** – The results highlight three findings. First, a fundamental link is confirmed between the two cognitive styles. Second, the relationship between field dependent users and holists is suggested to be more prominent than that of field independent users and serialists. Third, the interface design preferences of field dependent and field independent users can be split more clearly than those of holists and serialists.

**Originality/value** – The contributions of this study include a deeper understanding of the similarities and differences between field dependence/independence and holists/serialists as well as proposing a novel methodology for data analysis.

**Keywords** Worldwide web, Search engines, User studies, Information searches, Individual psychology  
**Paper type** Research paper



Journal of Documentation  
Vol. 66, No. 4, 2010

RESUMO EMPÍRICO

# Mental models of the bibliographic universe. Part 1: mental models of descriptions

Jan Pisanski and Maja Žumer

*Department of Library and Information Science and Book Studies,  
University of Ljubljana, Ljubljana, Slovenia*

Mental models.  
Part 1

643

Received 16 June 2009  
Revised 16 December 2009  
Accepted 16 December 2009

## Abstract

**Purpose** – The paper aims to present the results of the first two tasks of a user study looking into mental models of the bibliographic universe and especially their comparison to the Functional Requirements for Bibliographic Records (FRBR) conceptual model, which has not yet been user tested.

**Design/methodology/approach** – The paper employs a combination of techniques for eliciting mental models and consisted of three tasks, two of which, card sorting and concept mapping, are presented herein. Its participants were 30 individuals residing in the general area of Ljubljana, Slovenia.

**Findings** – Cumulative results of concept mapping show a strong resemblance to FRBR. Card sorts did not produce conclusive results. In both tasks, participants paid special attention to the original expression, indicating that a special place for it should be considered.

**Research limitations/implications** – The study was performed using a relatively small sample of participants living in a geographically limited space using relatively straight-forward examples.

**Practical implications** – Some solid evidence is provided for adoption of FRBR as the conceptual basis for cataloguing.

**Originality/value** – This is the first widely published user study of FRBR, applying novel methodological approaches in the field of Library and Information Science.

**Keywords** Bibliographics, User studies, Cataloguing, Cognition, Slovenia

**Paper type** Research paper

RESUMO EMPÍRICO

JDOC  
66,5

# Mental models of the bibliographic universe. Part 2: comparison task and conclusions

Jan Pisanski and Maja Žumer

*Department of Library and Information Science and Book Studies,  
University of Ljubljana, Ljubljana, Slovenia*

668

Received 16 June 2009  
Revised 16 December 2009  
Accepted 16 December 2009

## Abstract

**Purpose** – The paper aims to provide some insight into mental models of the bibliographic universe and how they compare with functional requirements for bibliographic records (FRBR) as a conceptual model of the bibliographic universe.

**Design/methodology/approach** – To get a more complete picture of the mental models, different elicitation techniques were used. The three tasks of the paper were: card-sorting, concept mapping and comparison task. The paper deals with comparison task, which consisted of interviews and rankings, and provides a discussion of the results of the paper as a whole.

**Findings** – Results of the ranking part of the comparison task confirm the findings of concept mapping task. In both cases, while there are individual differences between mental models, on average they gravitate towards FRBR.

**Research limitations/implications** – This is a small study and it provides only a glimpse of the implications of using FRBR as a conceptual basis for cataloguing. More FRBR-related user studies are needed, including similar studies on different groups of individuals and different types of materials, as well as practical studies of user needs and user interfaces.

**Practical implications** – The results of this study are the first user-tested indication of the validity of FRBR as a conceptual basis for the future of cataloguing.

**Originality/value** – This is the first published paper of mental models of the bibliographic universe and uses a unique combination of mental model elicitation techniques.

RESUMO EMPÍRICO

# Information literacy and knowledge management at work

## Conceptions of effective information use at NHS24

Rubén Toledano O'Farrill  
*Department of Economy, Management and Marketing,  
ITESO University, Guadalajara, Mexico*

### Abstract

**Purpose** – The purpose of this paper is to explore theoretically and empirically the concept of workplace information literacy (IL) and its connections to knowledge management (KM), and to examine the applicability of the mainstream institutional IL frameworks (American Library Association, The Australian and New Zealand Institute for IL, The Society of College, National and University Libraries, The UK Chartered Institute of Library Professionals, etc.) to the workplace environment.

**Design/methodology/approach** – Phenomenographic study of conceptions of effective information use of frontline staff at NHS24, which are discussed with respect to organizational characteristics of NHS Scotland and against the background of the two theoretical concepts: KM and IL.

**Findings** – The paper shows that the main institutional IL frameworks do not properly reflect some important ways in which information is used by participants in the study, particularly the use of people as information sources and the social sense making and interpretation of the value of information and its application in the workplace environment.

**Research limitations/implications** – The results from this qualitative study of a small population can be useful inasmuch as they highlight characteristics of workplace IL that may be relevant in several organizational settings, as well as presenting a critical analysis of the IL frameworks developed by librarians in regard to their applicability to workplace settings.

**Practical implications** – The paper highlights the importance of developing organization-specific guidelines for IL development, endorsing views of learning and information use grounded in socio-constructive perspectives and a consideration of context as situated practice.

**Originality/value** – This is an original research study conducted as part of a PhD qualification which has contributed to further the understanding of workplace IL.

**Keywords** Information literacy, Knowledge management, Workplace learning, Phenomenology,

RESUMO EMPÍRICO

# Engineering an image-browsing environment: re-purposing existing denotative descriptors

JungWon Yoon  
*School of Library and Information Science,  
University of South Florida, Tampa, Florida, USA, and*  
Brian O'Connor  
*Department of Library and Information Science, College of Information,  
University of North Texas, Denton, Texas, USA*

### Abstract

**Purpose** – The paper provides a theory base for deriving connotative descriptors for photographs from existing denotative descriptors, and then demonstrates a model for enhancing browsing within image collections by providing a tool for carving up the searching space.

**Design/methodology/approach** – The paper conceptually explores the nature of iconic messages contained in an image by adopting semiotics as a theoretical tool. A problem of image retrieval is identified as loss of connotative messages during the image representation process. The paper proposes an image-retrieval model utilizing an association thesaurus that facilitates the assignment of connotative index terms by making use of denotative index terms of an image. A series of experiments are performed for evaluating the effectiveness of the proposed model.

**Findings** – Experimental results demonstrate that the association thesaurus improves image-retrieval effectiveness by increasing the recall of connotatively related image documents as well as the recall of browsing sets.

**Practical implications** – Applying connotative index terms to an image would be time consuming. Deriving connotative terms from denotative terms and then using them to enrich the browsing environment suggest a method of increasing retrieval effectiveness while reducing the resources required for representation.

**Originality/value** – Since images are often used to illustrate concepts that are not immediately evident from just the objects in front of the lens, connotative descriptors are particularly valuable. Since human perception of images is, in a sense, hard wired into our brains, browsing is a frequent and reasonable search method in image collections. Using connotative descriptors to point the way to clusters of images with a higher probability of relevance changes the locus of control over representation and establishes an environment for dynamic representation, and gives credibility to browsing as a significant search method.

**Keywords** Photographs, Document image processing

**Paper type** Research paper

### Introduction

Browsing is a significant method for seeking documents (O'Connor, 1993). This is especially the case when there is some form of discrepancy or indeterminacy between the system of representation and the individual seeker's representation of his or her need.



RESUMO EMPÍRICO

## Documentary tools in everyday life: the wedding planner

Pamela J. McKenzie

*Faculty of Information and Media Studies, The University of Western Ontario,  
London, Canada, and*

Elisabeth Davies

*School of Information Management, Dalhousie University, Halifax, Canada*

788

Received 5 October 2009  
Revised 4 February 2010  
Accepted 11 February 2010

### Abstract

**Purpose** – This paper aims to analyze documentary planning tools for an everyday life project, the wedding, to study how “document work” is constructed in this setting.

**Design/methodology/approach** – Using Law and Lynch’s study of birdwatching guides for novices as a framework, nine commercially-available wedding planning guides targeted toward the primary planner, almost universally the bride, were analyzed.

**Findings** – As Law and Lynch found, part of a novice’s apprenticeship requires learning how to “see” in ways that are socially organized in and through texts. The paper shows how characteristics of birdwatching guides (naturalistic accountability, a picture theory of representation, and the strategic use of texts) are also evident in wedding planners, and how the very features that make these guides usable also occasion troubles for their users. Wedding planning guides treat the bride as a novice and instruct her in seeing wedding-related tasks and times as amenable to management. However, planning a wedding requires multiple tasks and times that may be intertwined in ways that make both their representation and their execution highly complex.

**Research limitations/implications** – The need for both temporal and thematic access highlights more general problems of knowledge organization in presenting a complex planning project in a linear and paper format.

**Originality/value** – As workplace principles of time and project management are increasingly applied to everyday life, this paper provides a needed case study of the ways that everyday recordkeeping contributes to the novice bride’s gendered apprenticeship and embeds her work within broader organizational and ideological systems.

**Keywords** Records management, Information management, Project planning, Time-based management, Marriage

Paper type Research paper

RESUMO EMPÍRICO

## Opening and closing rituals of the virtual reference service of the Internet Public Library

Jung-ran Park, Guisu Li and Amy Burger

*College of Information Science and Technology, Drexel University,  
Philadelphia, Pennsylvania, USA*

Opening and  
closing rituals  
of the IPL

807

Received 23 November 2009  
Revised 25 January 2010  
Accepted 11 February 2010

### Abstract

**Purpose** – The purpose of this paper is to explore the communicative rituals of opening and closing manifested in e-mail-based Internet Public Library’s (IPL’s) online reference interaction.

**Design/methodology/approach** – A total of 400 transcripts comprising user queries and responses by IPL librarians are examined. The opening and closing elements are identified to examine the way in which IPL librarians and users construct social space; that is, communicate their interpersonal and affective stances during the course of seeking and offering information.

**Findings** – The results of data analysis show regular patterns of verbal and structural politeness indicators of opening and closing e-mail discourse. Linguistic elements such as greetings and acknowledgement are included in all the sampled transcripts; i.e. a 100 percent occurrence. Closing rituals have a 95 percent occurrence of linguistic features such as acknowledgement and invitation for follow-up. In contrast, there is a low occurrence of personalized openings through greeting by user name (26 percent). This lack of personalization also occurs in closings: personalized farewell through use of librarian name appears in only 8 percent of closings.

**Research limitations/implications** – The employment of the various politeness tactics in opening and closing reflects the librarian’s attention and concern to user’s information needs, interests and wants. Such communicative competence narrows social distance and brings forth close socio-interpersonal space for interaction; this may, in turn, improve the overall quality of reference service. Research findings also indicate that more use of personal names may decrease the social distance between the librarian and user, resulting in increased solidarity and proximity.

**Originality/value** – The study provides new insights into linguistic politeness and the functions of address forms such as personal names with a view toward developing effective opening and closing rituals that contribute to the enhancement of virtual reference services.

**Keywords** Interpersonal communications, Internet, Public libraries, Virtual organizations, Reference services

Paper type Research paper

RESUMO EMPÍRICO

## The individual and social dynamics of knowledge sharing: an exploratory study

Sylvio Cyr  
*Meta RIM Consultants Ltd, Hong Kong, and*  
Chun Wei Choo  
*Faculty of Information, University of Toronto, Toronto, Canada*

### Abstract

**Purpose** – This paper aims to examine how knowledge sharing behavior is influenced by three sets of dynamics: a rational calculus that weighs the costs and benefits of sharing; a dispositional preference that favors certain patterns of sharing outcomes; and a relational effect based on working relationships.

**Design/methodology/approach** – Concepts from social exchange theory, social value orientation, and leader-member exchange theory are applied to analyze behavioral intentions to share knowledge. The study population consists of employees of a large pension fund in Canada. Participants answered a survey that used allocation games and situational vignettes to measure social value orientation, propensity to share knowledge, and perception of cost and benefit.

**Findings** – The results suggest that personal preferences about the distribution of sharing outcomes, individual perceptions about costs and benefits, and structural relationship with knowledge recipients, all affect knowledge sharing behavior significantly. Notably, it was found that propensity to share knowledge is positively related to perceived benefit to the recipient, thus suggesting that evaluation of cost and benefit in social exchange is not limited to self-interest, but is also influenced by perceived recipient benefit. Moreover, it was found that the relationship with the sharing target (superior or colleague) also influenced sharing.

**Originality/value** – Most studies emphasize the organizational benefits of knowledge sharing. This study examines knowledge sharing from the perspective of the individual who approaches knowledge sharing as a social exchange that involves perceptions of costs and benefits, preferences about sharing outcomes, and relationship with the sharing target. The study also introduces innovative methods to measure social value orientation and propensity to share knowledge.

**Keywords** Knowledge sharing, Social dynamics, Value analysis, Canada

**Paper type** Research paper

RESUMO EMPÍRICO

## Managing documents at home for serious leisure: a case study of the hobby of gourmet cooking

Jenna Hartel  
*Faculty of Information, University of Toronto, Toronto, Canada*

Managing documents at home

847

Received 28 August 2009  
Revised 18 February 2010  
Accepted 24 February 2010

### Abstract

**Purpose** – This paper aims to describe the way participants in the hobby of gourmet cooking in the USA manage culinary information in their homes.

**Design/methodology/approach** – The study utilizes domain analysis and serious leisure as a conceptual framework and employs an ethnographic approach. In total 20 gourmet cooks in the USA were interviewed at home and then their culinary information collections were documented through a guided tour and photographic inventory. The resulting ethnographic record was analyzed using grounded theory and NVivo software.

**Findings** – The findings introduce the personal culinary library (PCL): a constellation of cooking-related information resources and information structures in the home of the gourmet cook, and an associated set of upkeep activities that increase with the collection's size. PCLs are shown to vary in content, scale, distribution in space, and their role in the hobby. The personal libraries are characterized as small, medium or large and case studies of each extreme are presented. Larger PCLs are cast as a bibliographic pyramid distributed throughout the home in the form of a mother lode, zone, recipe collection, and binder.

**Practical implications** – Insights are provided into three areas: scientific ethnography as a methodology; a theory of documents in the hobby; and the changing role of information professionals given the increasing prevalence of home-based information collections.

**Originality/value** – This project provides an original conceptual framework and research method for the study of information in personal spaces such as the home, and describes information phenomena in a popular, serious leisure, hobby setting.

**Keywords** Information management, Document management, Leisure activities, Ethnography, United States of America

**Paper type** Research paper

### Introduction

RESUMO EMPÍRICO

# A multilevel model of HIV/AIDS information/help network development

A model of information/help development

875

Tiffany Veinot

*School of Information and School of Public Health,  
Department of Health Behavior and Health Education, University of Michigan,  
Ann Arbor, Michigan, USA*

Received 30 November 2009  
Revised 3 January 2010  
24 February 2010  
Accepted 28 February 2010

## Abstract

**Purpose** – This paper aims to describe the personal information and help networks of people with HIV/AIDS (PHAs) in rural Canada, and to present a research-based model of how and why these networks developed. This model seeks to consider the roles of PHAs, their family members/friends and formal health systems in network formation.

**Design/methodology/approach** – In-depth, semi-structured interviews were conducted with 114 PHAs, their friends/family members (FFs) and formal caregivers in three rural regions of Canada. A network solicitation procedure elicited PHAs' HIV/AIDS information/help networks. Interviews were analyzed qualitatively, and network data were analyzed statistically. Documents describing health systems in each region were also analyzed. Analyses used social capital theory, supplemented by stress/coping and stigma management theories.

**Findings** – PHAs' HIV/AIDS related information/help networks emphasized linking and bonding social capital with minimal bridging social capital. This paper presents a model that explains how and why such networks developed. The model shows that networks grew from the actions of PHAs, their FFs and health systems. PHAs experienced considerable stress, which led them to develop information/help networks to cope with HIV/AIDS – both individually and collaboratively. Because of stigmatization, many PHAs disclosed their illness selectively, thus constraining the size and composition of their networks. Health system actors created network-building opportunities for PHAs by providing them with care, referrals and support programs.

**Originality/value** – This study describes and explains an understudied type of information behavior: information/help network development at individual, group and institutional levels. As such, it illuminates the complex dynamics that made individual acts of interpersonal information acquisition and sharing possible.

**Keywords** Information management, Social networks, Stress, HIV, Acquired immune deficiency syndrome, Canada

RESUMO EMPÍRICO

JDOC  
66,6

# How poor informationally are the information poor?

906

## Evidence from an empirical study of daily and regular information practices of individuals

Liangzhi Yu

*Department of Information Resource Management, The Business School,  
Nankai University, Tianjin, People's Republic of China*

Received 19 December 2009  
Revised 9 March 2010  
Accepted 13 March 2010

## Abstract

**Purpose** – The purpose of this paper is to examine the multifaceted information disadvantages facing the so-called information poor in today's society. It has two specific objectives; to identify, from the empirical evidence of individuals' daily and regular information practices, meaningful constructs for defining information inequality; and to investigate how the "information poor" characterise on these constructs in comparison with others.

**Design/methodology/approach** – The study developed its findings in relation to the above objectives inductively from the interview transcripts of 73 people of different social statuses and occupations regarding their daily and regular information practices.

**Findings** – Three concepts emerged as meaningful constructs to define information inequality, hence to describe the information poor: individuals' information resource bases, information practices, and information assets. The information poor are found to be disadvantaged in all these aspects. They are first disadvantaged by limited freedom and/or opportunities in claiming society's vast and rapidly increasing information resources into their own information resource bases; then by the constraint of their information practices in developing their information resource bases and obtaining information utilities from these resources; and further more by impoverished information assets to empower themselves in normal and problematic situations.

**Practical implications** – Understanding of the information poor as sketched above will likely demand further research into a number of issues/areas hitherto ignored by information inequality studies.

**Originality/value** – This paper is one of the first to address empirically one of the most fundamental questions in information inequality studies; it extends library and information science understanding of the information poor hitherto dominated by the information deficiency thesis.

**Keywords** Information services, Information management, Disadvantaged groups, China  
Paper type Research paper

RESUMO EMPÍRICO

JDOC  
66,6

934

Received 15 January 2010  
Revised 11 June 2010  
Accepted 12 June 2010

## SPECULATIONS IN DOCUMENTATION Carbon footprint of the knowledge sector: what's the future?

Gobinda Chowdhury  
*Centre for Information and Knowledge Management,  
University of Technology, Sydney, Australia*

### Abstract

**Purpose** – The purpose of this paper is to produce figures showing the carbon footprint of the knowledge industry – from creation to distribution and use of knowledge, and to provide comparative figures for digital distribution and access.

**Design/methodology/approach** – An extensive literature search and environmental scan was conducted to produce data relating to the CO<sub>2</sub> emissions from various industries and activities such as book and journal production, photocopying activities, information technology and the internet. Other sources such as the International Energy Agency (IEA), Carbon Monitoring for Action (CARMA), Copyright Licensing Agency, UK (CLA), Copyright Agency Limited, Australia (CAL), etc., have been used to generate emission figures for production and distribution of print knowledge products versus digital distribution and access.

**Findings** – The current practices for production and distribution of printed knowledge products generate an enormous amount of CO<sub>2</sub>. It is estimated that the book industry in the UK and USA alone produces about 1.8 million tonnes and about 11.27 million tonnes of CO<sub>2</sub> respectively. CO<sub>2</sub> emission for the worldwide journal publishing industry is estimated to be about 12 million tonnes. It is shown that the production and distribution costs of digital knowledge products are negligible compared to the environmental costs of production and distribution of printed knowledge products.

**Practical implications** – Given the astounding emission figures for production and distribution of printed knowledge products, and the associated activities for access and distribution of these products, for example, emissions from photocopying activities permitted within the provisions of statutory licenses provided by agencies like CLA, CAL, etc., it is proposed that a digital distribution and access model is the way forward, and that such a system will be environmentally sustainable.

**Originality/value** – It is expected that the findings of this study will pave the way for further research and this paper will be extremely helpful for design and development of the future knowledge distribution and access systems.

**Keywords** Digital libraries, Environmental management, Global warming, Information management, Generation and dissemination of information

**Paper type** Research paper



Journal of Documentation

RESUMO EMPÍRICO

## Comedy as freedom of expression

Paul Sturges  
*Department of Information Science, Loughborough University,  
Loughborough, UK*

Comedy as  
freedom of  
expression

279

### Abstract

**Purpose** – The purpose of this paper is to investigate the practice of comedians in relation to freedom of expression, so as to throw light on the issue of giving or avoiding offence.

**Design/methodology/approach** – The literature of comedy, newspaper coverage of comedy in the UK in 2008, observation of comedians in performance, and a small, informal interview programme with stand up comedians were used in the preparation of the paper.

**Findings** – Stand up comedians, despite their own sense that they defy restriction and popular perception of their material as often offensive, do monitor their material for potential offence. They assess the extent of offence and modify their performances in response. In some cases they apply personal formulae to this process.

**Research limitations/implications** – The interview programme is too small to claim to be fully representative and is intended only to give an indicative view of the field.

**Practical implications** – Examination of comedians' practice has implications for information service institutions and the giving of access to potentially offensive content.

**Originality/value** – The paper may be the first study of comedy in an information science context and it contains implications for further studies that use comedy as an example of content, and creative practice to further develop understanding of information provision issues.

**Keywords** United Kingdom, Information exchange, Entertainment industry, Freedom of information, Humour

**Paper type** Research paper

Received 4 March 2009  
Revised 27 July 2009  
Accepted 29 July 2009

RESUMO EMPÍRICO

## The Foundation of the Concept of Relevance

Birger Hjørland

*The Royal School of Library and Information Science, 6 Birketinget, DK-2300  
Copenhagen S, Denmark. E-mail: bh@db.dk*

In 1975 Tefko Saracevic declared "the subject knowledge view" to be the most fundamental perspective of relevance. This paper examines the assumptions in different views of relevance, including "the system's view" and "the user's view" and offers a reinterpretation of these views. The paper finds that what was regarded as the most fundamental view by Saracevic in 1975 has not since been considered (with very few exceptions). Other views, which are based on less fruitful assumptions, have dominated the discourse on relevance in information retrieval and information science. Many authors have reexamined the concept of relevance in information science, but have neglected the subject knowledge view, hence basic theoretical assumptions seem not to have been properly addressed. It is as urgent now as it was in 1975 seriously to consider "the subject knowledge view" of relevance (which may also be termed "the epistemological view"). The concept of relevance, like other basic concepts, is influenced by overall approaches to information science, such as the cognitive view and the domain-analytic view. There is today a trend toward a social paradigm for information science. This paper offers an understanding of relevance from such a social point of view.

### Introduction

In 1975 Tefko Saracevic<sup>1</sup> declared "the subject knowledge view" to be the most fundamental perspective of relevance: "I wish to suggest that the *subject knowledge view of relevance*

We have the strange situation that what is seen as the most fundamental perspective has not been considered further.

If Saracevic was right that "the subject knowledge view" is the most fundamental perspective of relevance, it follows that the competing perspectives on relevance have (implicit) defects—at least I shall argue so. However, these defects have not hindered their diffusion and popularity in information science.

This paper contains arguments that the dominant views on relevance today are less fruitful than the subject knowledge view. I consider this issue related to the developments of metatheoretical discourses such as the cognitive view and the domain-analytic view in information science. The received view is that the user-oriented view of relevance became dominant in the second half of the 1970s. Before that "the systems view of relevance" was the dominant view. I have argued elsewhere (e.g., Hjørland & Albrechtsen, 1995; Hjørland, 2009) that a social paradigm is a better theoretical framework than the cognitive view.<sup>2</sup> This has also been argued by Cronin (2008). This paper develops a view of relevance informed by a social paradigm (termed domain-analysis or the socio-cognitive paradigm), which corresponds with what Saracevic termed "the subject knowledge view." It constitutes the first full-length article attempting to situate "the subject knowledge view" as a foundation for understanding the concept of relevance within information retrieval/information science

RESUMO DISSERTATIVO

## Referencing in the Humanities and its Implications for Citation Analysis

Björn Hellqvist

*Uppsala Universitet, Institution of ALM (Archives, Libraries & Museums), Box 625,  
751 26 Uppsala, Sweden. E-mail: Bjorn.hellqvist@abm.uu.se*

This article studies citation practices in the arts and humanities from a theoretical and conceptual viewpoint, drawing on studies from fields like linguistics, history, library & information science, and the sociology of science. The use of references in the humanities is discussed in connection with the growing interest in the possibilities of applying citation analysis to humanistic disciplines. The study shows how the use of references within the humanities is connected to concepts of originality, to intellectual organization, and to searching and writing. Finally, it is acknowledged that the use of references is connected to stylistic, epistemological, and organizational differences, and these differences must be taken into account when applying citation analysis to humanistic disciplines.

### Introduction

The aim of this article is to study the meaning and function of referencing within the arts and humanities in order to

disciplines in the humanities, where little research concerning publication and citing practices has been done.

When discussing the possibility of using citation analysis to evaluate humanistic disciplines the problems discussed are often of a technical nature: databases such as the Web of Science are too narrow in scope, humanistic scholars publish in their native language and not in English-language journals, and they publish in monographs and anthologies rather than journals. These are obviously serious problems that have to be dealt with when trying to study and evaluate the humanities using citation analysis. However, an epistemological aspect of equal importance is how and why you cite. Citation databases and methods do not sufficiently cover the humanities but neither do citation theories developed to explain the use of citations in science. This article does not propose a 'citation theory' for the humanities but it tries to depict important characteristics of referencing in the humanities that affect the application of citation analysis.

RESUMO DISSERTATIVO

## Requirements for the Design of a Personal Document-Management System

**Olha Bondarenko**

*Eindhoven University of Technology, 5600MB Eindhoven, The Netherlands. E-mail: o.bondarenko@tue.nl*

**Ruud Janssen**

*Novay, 7500AN Enschede, The Netherlands. E-mail: ruud.janssen@novay.nl*

**Samuël Driessen**

*Océ-Technologies, 5900MA Venlo, The Netherlands. E-mail: samuel.driessen@oce.com*

In this article a set of requirements for the design of a personal document management system is presented, based on the results of three research studies (Bondarenko, 2006; Bondarenko & Janssen, 2005; Bondarenko & Janssen, 2009). We propose a framework, based on layers of task decomposition, that helps to understand the needs of information workers with regard to personal document and task management. Relevant user processes are described and requirements for a document-management system are derived for each layer. The derived requirements are compared to related studies, and implications for system design are discussed.

instant messaging tools, forums, blogs, chat rooms, social-networking tools, Web conferencing, and voice-over-IP. Third, the number of devices and tools that an information worker has to operate has also increased dramatically. At the beginning of the 21st century, the physical workplace of an information worker has acquired its digital counterpart, currently consisting of a variety of devices, from PCs to handheld computers and mobile phones. There is an even bigger variety of software tools, from e-mail clients and office software to dedicated applications for professional use. New devices and new versions of software are frequently released, forcing users to learn tools and acquire new skills, while the number

### RESUMO EXPOSITIVO

## A Robust Benchmark for the h- and g-Indexes

**Giovanni Abramo**

*National Research Council of Italy and Dipartimento di Ingegneria dell'Impresa, University of Rome "Tor Vergata," Via del Politecnico 1, 00133 Rome, Italy. E-mail: abramo@disp.uniroma2.it*

**Ciriaco Andrea D'Angelo and Fulvio Viel**

*Dipartimento di Ingegneria dell'Impresa, University of Rome "Tor Vergata," Via del Politecnico 1, 00133 Rome, Italy.*

The use of Hirsch's h-index as a joint proxy of the impact and productivity of a scientist's research work continues to gain ground, accompanied by the efforts of bibliometrists to resolve some of its critical issues through the application of a number of more or less sophisticated variants. However, the literature does not reveal any appreciable attempt to overcome the objective problems of measuring h-indexes on a large scale for purposes of comparative evaluation. Scientists may succeed in calculating their own h-indexes but, being unable to compare them to those of their peers, they are unable to obtain truly useful indications of their individual research performance. This study proposes to overcome this gap, measuring the h- and Egge's g-indexes of all Italian university researchers in the hard sciences over a 5-year window. Descriptive statistics are provided concerning all of the 165 subject fields examined, offering robust benchmarks for those who wish to compare their individual performance to those of their colleagues in the same subject field.

#### Introduction

The h-index was first introduced as a "measure" of the accomplishments of an entire scientific career.<sup>1</sup> However,

ment to a population of scientists in a subject field sufficiently numerous to offer a robust benchmark. Otherwise, there is no explanation of why there are still no statistics for the h-index by subject field, as there are for other measures such as average citations per article, where Thomson Scientific calculates the pertinent data for each subject category. Just as for an elite runner, scientists will not be content to know only their own results because the drive to improve is partially dependent on the opportunity to compare with one's peers. Such comparisons are even more essential for the decision maker who, to allocate resources in an efficient manner, must not only compare scientists within the same subject field, but also among different subject fields. The elite runner will also want to know if peers' results have been achieved under similar conditions (wind, altitude, etc.). In terms of scientists' production, the analogous need is one of knowing if the period of comparison is the same, if more than one scientist participated in the output (coauthorship), and if the resources used and the potential economic rents were the same. The numerous variants of the h-index (and variants of the variants) that follow one another in the literature represent attempts to perfect the indicator, rendering it as objective and representative of

### RESUMO EXPOSITIVO

# Problems of Citation Analysis: A Study of Uncited and Seldom-Cited Influences

M.H. MacRoberts and B.R. MacRoberts

*Herbarium, Museum of Life Sciences, Biology Department, Louisiana State University, Shreveport, LA 71115, and Bog Research, 740 Columbia, Shreveport, LA 71104. E-mail: mmacrobe@lsus.edu*

To determine influences on the production of a scientific article, the content of the article must be studied. We examined articles in biogeography and found that most of the influence is not cited, specific types of articles that are influential are cited while other types of that also are influential are not cited, and work that is "uncited" and "seldom cited" is used extensively. As a result, evaluative citation analysis should take uncited work into account.

## Introduction

In a previous series of studies, we examined articles and compared influences evident in the text with those referenced in the bibliography. Regardless of subject examined, we found that authors cite only a fraction of their influence

2006). The main question asked by biogeographers is why are plants and animals distributed as they are, and thus "why are there this many species in any particular locality?" Major sources of information for plants are herbarium specimens and floras; main sources for zoogeography are museum collections of animals and "faunas" (The word is not currently in general use, but since there is no general term today for these works, we use it here.). Authors of biogeographical articles have not personally collected the data they use but rely heavily on extremely large databases, compiled by thousands of individuals over centuries. However, what is particularly interesting about articles in this field from the standpoint of citation analysis is that there is a generally accepted protocol

## RESUMO EMPÍRICO

### Consumer Health Information on the Web: The Relationship of Visual Design and Perceptions of Credibility

David Robins

*Kent State University, Information Architecture and Knowledge Management, Kent, OH.  
E-mail: drobins@kent.edu*

Jason Holmes

*American Greetings Interactive, Cleveland, OH.*

Mary Stansbury

*University of Denver, Department of Library and Information Science, Denver, CO.*

Consumer health information has proliferated on the Web. However, because virtually anyone can publish this type of information on the Web, consumers cannot always rely on traditional credibility cues such as reputation of a journal. Instead, they must rely on a variety of cues, including visual presentation, to determine the veracity of information. This study is an examination of the relationship of people's visual design preferences to judgments of credibility of information on consumer health information sites. Subjects were asked to rate their preferences for visual designs of 31 health information sites after a very brief viewing. The sites were then reordered and subjects rated them according to the extent to which they thought the information on the sites was credible. Visual design judgments bore a statistically significant similarity to credibility ratings. Sites with known brands were also highly rated for both credibility and visual design. Theoretical implications are discussed.

is critical that information provided for consumers be accurate to avoid negative consequences of bad diagnoses or advice.

In October of 2006, a Pew Charitable Trust survey found that 8 of 10 respondents use the Web for health research (Fox, 2006). Of those, 48%–56% searched on behalf of another person. Fifty-eight percent of the respondents said that the information found on a Web site had an effect on a health care decision.

In this article, we lean heavily on two particular terms: visual design and credibility. Visual design is the rendering of information, in the case of this study, on a Web site. It includes structural features such as typography, images, color, and aesthetics. Visual design includes grid systems that divide pages into different sectors into which content is placed. Visual design is not so much about *what* is communicated but rather about *how* information is communicated

## RESUMO EMPÍRICO

## The Development and Evaluation of a Survey to Measure User Engagement

Heather L. O'Brien

School of Library, Archival and Information Studies, University of British Columbia, Vancouver, BC V6T 1Z1 Canada. E-mail: hlobrien@interchange.ubc.ca

Elaine G. Toms

School of Business Administration, Dalhousie University, Halifax, Nova Scotia, B3H 3J5 Canada. E-mail: Elaine.Toms@dal.ca

Facilitating engaging user experiences is essential in the design of interactive systems. To accomplish this, it is necessary to understand the composition of this construct and how to evaluate it. Building on previous work that posited a theory of engagement and identified a core set of attributes that operationalized this construct, we constructed and evaluated a multidimensional scale to measure user engagement. In this paper we describe the development of the scale, as well as two large-scale studies (N=440 and N=802) that were undertaken to assess its reliability and validity in online shopping environments. In the first we used Reliability Analysis and Exploratory Factor Analysis to identify six attributes of engagement: Perceived Usability, Aesthetics, Focused Attention, Felt Involvement, Novelty, and Endurability. In the second we tested the validity of and relationships among those attributes using Structural Equation Modeling. The result of this research is a multidimensional scale that may be used to test the engagement of software applications. In addition, findings indicate that attributes of engagement are highly intertwined, a complex interplay of user-system interaction variables. Notably, Perceived Usability played a mediating role in the relationship between Endurability and Novelty, Aesthetics, Felt Involvement, and Focused Attention.

making up their minds about what technologies to invest their time, effort, and dollars in based on how they make them feel. Thus, the question is no longer only whether an application is efficient, effective, or satisfying, but how well it is able to engage users and provide them with an experience (Bannon, 2005; Overbeeke et al., 2003). The experience movement is being driven both by the desire to identify system features that facilitate positive user experiences and apply them to the design of technologies (Lingaard, 2004), and to market beyond tangible goods and services, and incorporate experience as a means of differentiating goods and services in a competitive environment (Pine & Gilmour, 1999).

Designing for engaging experiences is an oft-cited goal of interactive system development in many disciplines, yet there are no guidelines to channel designers' efforts to "make things engaging" (Overbeeke et al., 2003). Part of the issue has been that, until recently, there have been disparate ideas about what constitutes engagement. Without a consistent definition of engagement, it is difficult to ascertain that the systems we design and market are, in fact, engaging, or to identify what aspects of the interaction with technology engage or fail to engage users. Without a means of measuring engagement, we are limited in our understanding of this quality of user

### RESUMO EMPÍRICO

## Analysis of Factors Influencing Application of ICT by Agricultural Graduate Students

Mahtab Pouratashi and Ahmad Rezvanfar

Department of Agricultural Extension and Education, Faculty of Agricultural Economics and Development, University of Tehran, Karaj, Iran. E-mail: m\_pouratashi@yahoo.com

With improving technology in the world, particularly Information and Communication Technology (ICT), people need to make an effort to keep up with the new technology. The importance of ICT in higher education continues to grow year by year. The purpose of this study was to analyze factors influencing application of ICT by agricultural graduate students. The statistical population included agricultural graduate students in colleges of agriculture at the University of Tehran. A sample of 110 students was selected using a random sampling method. A questionnaire was used for data collection. Reliability and validity of instrument were determined through opinions of faculty members and application of Cronbach's alpha. The data were analyzed using descriptive and inferential statistics. The findings revealed that skill, support, and facilities were the three factors that influenced the application of ICT by agricultural students. In order to predict application of ICT by agricultural students, the Technology Acceptance Model (TAM) was used. The results showed that skill had direct and indirect effects on the application of ICT, while support and facilities affected the application of ICT indirectly. Given the direct effect on application of ICT, we infer that when students' skills improve, they are more likely to use ICT.

vital role in the lives of individuals. Information is a powerful tool in addressing agricultural needs and if it is used appropriately it can radically change a nation's economy (Tshabalala, 2001).

There has been rapid growth and improvement in the diffusion of technologies that have been designed to handle information and aid communication. Information and Communication Technologies (ICTs) such as mobile devices and Web browsers are becoming increasingly important in people's daily lives (Ye, Seo, Desouza, Sangareddy, & Jha, 2008). ICTs power our access to information (Brown, 2002), offer new ways of communication (Brown, 2002; Nachmias, Mioduser, & Shemla, 2001), and serve many online services in the spheres of commerce, entertainment, and education (Nachmias et al., 2001).

Education can be hypothesized to be one of the key propellants for national and regional prosperity, especially in developing countries. A key role of education is to empower students with attitudes and skills that are essential to their success in information society. ICTs became the essential factor for educational development in the current century (Kozma & Anderson, 2002; Goodison, 2003; Kangro & Kangro, 2004; Hennessy, Ruthven, & Brindley, 2005; Mobaideen, 2006; Kennedy, Krause, Judd, Churchward, & Gray, 2006; Bayindir & Inan, 2009). ICT supports teach-

### Introduction

The agricultural sector is facing a range of challenges

### RESUMO EMPÍRICO

## On the Relationship Between Interdisciplinarity and Scientific Impact

Vincent Larivière and Yves Gingras

Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, CP 8888, Succursale Centre-ville, Montréal, Québec, H3C 3P8 Canada. E-mail: {lariviere.vincent, gingras.yves}@uqam.ca

This article analyzes the effect of interdisciplinarity on the scientific impact of individual articles. Using all the articles published in *Web of Science* in 2000, we define the degree of interdisciplinarity of a given article as the percentage of its cited references made to journals of other disciplines. We show that although for all disciplines combined there is no clear correlation between the level of interdisciplinarity of articles and their citation rates, there are nonetheless some disciplines in which a higher level of interdisciplinarity is related to a higher citation rates. For other disciplines, citations decline as interdisciplinarity grows. One characteristic is visible in all disciplines: Highly disciplinary and highly interdisciplinary articles have a low scientific impact. This suggests that there might be an optimum of interdisciplinarity beyond which the research is too dispersed to find its niche and under which it is too mainstream to have high impact. Finally, the relationship between interdisciplinarity and scientific impact is highly determined by the citation characteristics of the disciplines involved: Articles citing citation-intensive disciplines are more likely to be cited by those disciplines and, hence, obtain higher citation scores than would articles citing non-citation-intensive disciplines.

### Introduction

Over the last 40 years and especially since the pub-

has been analyzed, especially in the context of the British Research Assessment Exercise—where the question whether researchers involved in interdisciplinary research should be evaluated differently from those doing disciplinary research was raised—these studies are usually based on a small sample of disciplines, university departments, or journals. For instance, measuring interdisciplinarity at the level of journals, Levitt and Thelwall (2008) found that in the natural and health sciences, multidisciplinary research (defined as articles published in journals to which more than one discipline has been attributed) have less scientific impact than does disciplinary research. In the social sciences, both types of research obtain similar citation rates. Using the interdisciplinarity of cited references among Thomson *Current Contents* 88 subject categories and discipline-normalized citation counts of two research-intensive United Kingdom universities ( $N = 37,000$  articles), Adams, Jackson, and Marshall (2007) found that the most interdisciplinary articles were in fact cited as much as the average article. They also found that the cited references of the most cited articles had average levels of interdisciplinarity. Finally, using physics research programs in the Netherlands as a case study, Rinia, van Leeuwen, and van Raan (2002) showed that absolute bibliometric indicators are indeed biased against interdisciplinary research.

RESUMO EMPÍRICO

## Testing the Trade-off Between Productivity and Quality in Research Activities

Giovanni Abramo

National Research Council of Italy and Laboratory for Studies of Research and Technology Transfer, Dipartimento di Ingegneria dell'Impresa, Università degli Studi di Roma Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy. E-mail: abramo@disp.uniroma2.it

Ciriaco Andrea D'Angelo

Laboratory for Studies of Research and Technology Transfer, Dipartimento di Ingegneria dell'Impresa, Università degli Studi di Roma Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy. E-mail: dangelo@disp.uniroma2.it

Flavia Di Costa

Laboratory for Studies of Research and Technology Transfer, Centro Interdipartimentale Vito Volterra, Università degli Studi di Roma Tor Vergata, Via Columbia 2, 00133 Rome, Italy. E-mail: dcosta@disp.uniroma2.it

In recent years there has been an increasingly pressing need for the evaluation of results from public-sector research activity, particularly to permit the efficient allocation of scarce resources. Many of the studies and evaluation exercises that have been conducted at the national and international levels emphasize the quality dimension of research output, while neglecting that of productivity. This work is intended to test for the possible existence of correlation between quantity and quality of scientific production and determine whether the most productive researchers are also those that achieve results that are qualitatively better than those of their colleagues. The analysis proposed refers to the entire Italian university system and is based on the observation of production in the hard sciences by more than 26,000 researchers in the period 2001–2006. The results show that the output of more-productive researchers is superior in quality than that of less-productive researchers. The relation between productivity and quality results is largely insensitive to the type of indicators or the test methods applied and also seems to differ little among the various disciplines examined.

### Introduction

In recent years, the debate over the evaluation of publicly funded research has attracted interest from a wide and

varied public. More evaluation exercises and analyses are being implemented, for the examination of both public institutions and individual researchers. Comparative evaluation would allow stimulation of greater productive efficiency, permit allocation of resources as a function of merit, and reduce information asymmetry between demand and supply of knowledge.

However, comparative evaluation of research activity results is quite complex, especially along the dimension for labor productivity. The national research assessments implemented in various nations have so far taken a peer-review approach, avoiding the evaluation of research productivity and all of its associated difficulties. These peer-review assessments emphasize examination of quality in research output, and limit observations to a sample of total output provided by each institution under observation (for example, 10% in the case of the Valutazione Triennale della Ricerca [Triennial Research Assessment] [VTR], Italy [CNR, 2006], and four research outputs for each researcher over a 6-year period in the case of the United Kingdom [KAE, 2008]). This is usually a deliberate policy decision. However, it must be recognized that national-level measures of productivity are not possible with a peer-review approach unless the evaluation is extended to the entire scientific production of the institutions being evaluated. This would result in exercises with extremely high costs and long time spans for their execution. Meanwhile, the option of basing measures of productivity

Received July 30, 2009; revised September 15, 2009; accepted September 16, 2009

© 2009 ASIS&T • Published online 23 October 2009 in Wiley InterScience

RESUMO EMPÍRICO

## Individual Differences in the Interpretation of Text: Implications for Information Science

Jane Morris

Faculty of Information, University of Toronto, Toronto, Ontario, Canada. E-mail: jane.morris@utoronto.ca

Many tasks in library and information science (e.g., indexing, abstracting, classification, and text analysis techniques such as discourse and content analysis) require text meaning interpretation, and, therefore, any individual differences in interpretation are relevant and should be considered, especially for applications in which these tasks are done automatically. This article investigates individual differences in the interpretation of one aspect of text meaning that is commonly used in such automatic applications: lexical cohesion and lexical semantic relations. Experiments with 26 participants indicate an approximately 40% difference in interpretation. In total, 79, 83, and 89 lexical chains (groups of semantically related words) were analyzed in 3 texts, respectively. A major implication of this result is the possibility of modeling individual differences for individual users. Further research is suggested for different types of texts and readers than those used here, as well as similar research for different aspects of text meaning.

Silber & McCoy, 2002), topic tracking (Hollingsworth & Toefel, 2005), text skimming (Hollingsworth & Toefel), spelling correction (Hirst & Budanitsky, 2004), indexing (Al-Halimi & Kazman, 1998), and information retrieval (Felbaum, 1998). What has not been considered, however, are individual differences in the interpretation of lexical cohesion in text. Furthermore, individual differences in the interpretation of other aspects of text meaning such as coherence, paragraph meaning, and text structure have also been ignored computationally.

This article briefly explores approaches to individual differences in text interpretation from the disciplines of literary theory and library and information science and in computation linguistics as evidenced by the "problem" of inter-annotator agreement. In library and information science, the much studied problem of inter-indexer consistency is also evidence of individual differences in the interpretation

### RESUMO EMPÍRICO

## Analyzing User Interaction with the ViewFinder Video Retrieval System

Dan Albertson

School of Library and Information Studies, University of Alabama, 513 Gorgas Library, Box 870252, Tuscaloosa, AL 35487-0252. E-mail: dea@ua.edu

This study investigates interactive video retrieval. The basis for this study is that user- and search task-centric research in video information retrieval can assist efforts for developing effective user interfaces and help complement the existing corpus of video retrieval research by providing evidence for the benefits of evaluating systems using such an approach. Accordingly, the results were collected and analyzed from the perspective of certain users and search tasks (i.e., information needs). The methodology of this study employed specially designed interactive search experiments to examine a number of different factors in a video retrieval context, including those that correspond to search tasks of a particular domain, interface features and functions, system effectiveness, and user interactions. The results indicated that the use and effectiveness of certain interface features and functions were dependent on the type of search task, while others were more consistent across the full experiment. Also included is a review of prior research pertaining to visual search tasks, systems development, and user interaction. ViewFinder, the prototype system used to carry out the interactive search experiments of this study, is fully described.

### Introduction

This study analyzes task-related factors impacting inter-

browse capabilities.<sup>1</sup> Visual searching, on the other hand, utilizes automatically extractable attributes, such as colors, shapes, and textures, from individual frames to retrieve video content.<sup>2</sup> While a significant number of video retrieval studies exclusively emphasize and evaluate performance of visual searching, results from studies employing standardized evaluation measures, including those participating in the TREC Video Retrieval Evaluation Workshop (TRECVID), regularly show that keyword search functions are more reliable and effective for retrieving video. Meanwhile, formulating practical foundations for designing user interfaces of video retrieval systems remains an increasingly challenging priority for research. Improving means for system evaluation and design will likely always present a compelling agenda for video retrieval research, and the following examination endeavors to complement the existing corpus of studies investigating interactive video retrieval.

### Nature of the Study

This study investigates interface designs for video retrieval systems. Its importance is supported by the notion that users require more specialized support, i.e., interface features and functions, when searching and browsing visual information

### RESUMO EMPÍRICO

## The Natural Landscape Metaphor in Information Visualization: The Role of Commonsense Geomorphology

Sara Irina Fabrikant  
Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.  
E-mail: sara@geouzh.ch

Daniel R. Montello  
Department of Geography, University of California at Santa Barbara, Santa Barbara, CA 93106.  
E-mail: montello@geog.ucsb.edu

David M. Mark  
Department of Geography, University at Buffalo (SUNY), Buffalo, NY 14261. E-mail: dmark@buffalo.edu

The landscape metaphor was one of the first methods used by the information visualization community to reorganize and depict document archives that are not inherently spatial. The motivation for the use of the landscape metaphor is that everyone intuitively understands landscapes. We critically examine the information visualization designer's ontologies for implementing spatialized landscapes with ontologies of the geographic domain held by lay people. In the second half of the article, we report on a qualitative study where we empirically assessed whether the landscape metaphor has explanatory power for users trying to make sense of spatialized views, and if so, in what ways. Specifically, we are interested in uncovering how lay people interpret hills and valleys in an information landscape, and whether their interpretation is congruent with the current scientific understanding of geomorphologic processes. Our empirical results suggest that neither developers' nor lay users' understanding of terrain visualizations is based on universal understanding of the true processes that has shaped a natural landscape into hills and valleys, mountains, and canyons. Our findings also suggest that the information landscape metaphor for sense making of a document collection is not self-evident to lay users, as claimed by information landscape designers. While a deep understanding of geomorphology will probably not be required to successfully use an information landscape, we do suggest that a coherent theory on how people use space will be necessary to produce cognitively useful information visualizations.

**Introduction**

With the advance of the World Wide Web and its diverse applications for the exchange of information, many have tried

to conceptualize and depict digital information collections such as online libraries, Web-based multimedia collections or digital discussion forums (Dodge & Kitchin, 2000). Information providers and users have considered properties of digital information collections represented as "spaces," including their size, shape, and form. What is the nature of an information space, and what do they look like when depicted? How is information located in these spaces, and how do their spatial properties represent nonspatial properties of the information? Do depictions of information spaces (i.e., external visualizations) correspond with the mental maps (i.e., internal visualizations) of the information spaces held by their users? Depicting information collections as concrete spatial layouts, even when the collections are not themselves explicitly spatial, is an information visualization technique known as *spatialization* (Kuhn & Blumenthal, 1996). Spatialization is based on utilizing spatial metaphors such as location, distance, size, and connectivity to devise graphic displays that depict the structure or content of nonspatial information stored in very large databases. Instead of presenting users with large tables or long lists of queried items in text format, information spatializations allow users to explore graphic displays of information, often allowing users to see the "layout" of the information in a single view (Okunin & Fabrikant, 2003). Spatialization attempts to exploit the power of graphic displays as cognitive aids, allowing people to sift more efficiently through and gain knowledge from vast amounts of accumulated data such as medical records, banking transactions, or information-bearing items stored in online archives (Fabrikant & Buttenfield, 2001). However, empirical findings have suggested that information landscapes should be based on sound geographical principles (Fabrikant, 2000, 2001a, 2001b) and adhere to cartographic design standards

Received March 9, 2009; revised August 13, 2009; accepted August 14, 2009

RESUMO EMPÍRICO

## Social Validation of Collective Annotations: Definition and Experiment

Guillaume Cabanac  
Université de Toulouse, IRIT UMR 5505 CNRS, 118 route de Narbonne, F-31062 Toulouse cedex 9, France.  
E-mail: Guillaume.Cabanac@irit.fr

Max Chevalier  
Université de Toulouse, IRIT UMR 5505 CNRS, 118 route de Narbonne, F-31062 Toulouse cedex 9, France, and Université de Toulouse, LGC EA 2043, IUT Paul Sabatier, 129 avenue de Rangueil, BP 67701, F-31077 Toulouse cedex 4, France. E-mail: Max.Chevalier@irit.fr

Claude Chrisment and Christine Julien  
Université de Toulouse, IRIT UMR 5505 CNRS, 118 route de Narbonne, F-31062 Toulouse cedex 9, France. E-mail: (Claude.Chrisment, Christine.Julien)@irit.fr

People taking part in argumentative debates through collective annotations face a highly cognitive task when trying to estimate the group's global opinion. In order to reduce this effort, we propose in this paper to model such debates prior to evaluating their "social validation." Computing the degree of global confirmation (or refutation) enables the identification of consensual (or controversial) debates. Readers as well as prominent information systems may thus benefit from this information. The accuracy of the social validation measure was tested through an online study conducted with 121 participants. We compared their human perception of consensus in argumentative debates with the results of the three proposed social validation algorithms. Their efficiency in synthesizing opinions was demonstrated by the fact that they achieved an accuracy of up to 84%.

Introduction and Motivations

With the widespread adoption of digital documents both in the workplace and at home, people have felt frustrated at not being able to annotate them (Sellen & Harper, 2003, p. 96). Such a need led both research labs and companies to design a plethora of annotation systems—mostly targeting Web documents—since the 1990s (Wolfe, 2002). At first, they implemented the usual paper-based annotation functions, mainly for personal use. Then, taking advantage of modern computer storage and networking capabilities, annotation systems provided novel features for collective use. In particular, they enabled annotation sharing through dedicated servers, so that users could view and access them in context, i.e., within the annotated document. Moreover, subsequent readers may later reply to any annotation or any reply in turn, thus forming a debate anchored to the discussed passage. Such a debate is also called a "discussion thread" (e.g.,

RESUMO EMPÍRICO

## Is Scientific Literature Subject to a 'Sell-By-Date'? A General Methodology to Analyze the 'Durability' of Scientific Documents

Rodrigo Costas, Thed N. van Leeuwen, and Anthony F.J. van Raan  
Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62A, P.O. Box 905,  
2300 AX Leiden, The Netherlands. E-mail: {rocostas,leeuwen,vanraan}@cwts.leidenuniv.nl

The study of the citation histories and ageing of documents are topics that have been addressed from several perspectives, especially in the analysis of documents with "delayed recognition" or "sleeping beauties." However, there is no general methodology that can be extensively applied for different time periods or research fields. In this article, a new methodology for the general analysis of the ageing and "durability" of scientific papers is presented. This methodology classifies documents into three general types: delayed documents, which receive the main part of their citations later than normal documents; flashes in the pan, which receive citations immediately after their publication but are not cited in the long term; and normal documents, documents with a typical distribution of citations over time. These three types of durability have been analyzed considering the whole population of documents in the Web of Science with at least 5 external citations (i.e., not considering self-citations). Several patterns related to the three types of durability have been found and the potential for further research of the developed methodology is discussed.

### Introduction

One topic of high concern of researchers is how many years they need to wait until their papers can be properly

business in which research results will prove their impact soon after publication. This attitude underestimates the crucial importance of durability of research results. Therefore, the study of the ageing and delayed reception of scientific documents is certainly also an important topic for the application of bibliometrics in research assessment, for instance, the correct establishment of citation windows and proper calculation of impact indicators.

The study of ageing and citation histories of documents has been addressed from several perspectives (Aversa, 1985; Glänzel & Schoepflin, 1995; Moed, van Leeuwen, & Reesdijk, 1998; Aksnes, 2003), especially the analysis and detection of documents with "delayed recognition" (Garfield, 1980; Glänzel, Schlemmer, & Thijs, 2003) or even "sleeping beauties" (van Raan, 2004). The opposite effect of delayed recognition was also described as "flash in the pan" (van Dalen & Henkens, 2005): documents that are noticed immediately after publication and frequently cited but do not seem to have a lasting impact and die early in life. Both aspects (delayed recognition and flashes in the pan) deal with the more general ideas on durability and obsolescence of documents, as well as with the ageing of scientific literature, which is strongly related to the impact of documents over time.

RESUMO EMPÍRICO

## Knowledge Diffusion Through Publications and Citations: A Case Study Using ESI-Fields as Unit of Diffusion

Yuxian Liu  
Reading Section, Library of Tongji University, Siping Street 1239, 200092 Shanghai,  
PR China, and University of Antwerp, IBW, Venusstraat 35, B-2000 Antwerpen, Belgium.  
E-mail: yxliu@lib.tongji.edu.cn

Ronald Rousseau  
KHBO (Association K.U.Leuven), Industrial Sciences and Technology, Zeedijk 101, B-8400 Oostende,  
Belgium, and University of Antwerp, IBW, Venusstraat 35, B-2000 Antwerpen, Belgium, and  
Katholieke Universiteit Leuven, Dept. Mathematics, Celestijnenlaan 200B, B-3000 Leuven (Heverlee),  
Belgium. E-mail: ronald.rousseau@khbo.be

Two forms of diffusion are studied: diffusion by publications, originating from the fact that a group publishes in different fields; and diffusion by citations, originating from the fact that the group's publications are cited in different fields. The first form of diffusion originates from an internal mechanism by which the group itself expands its own borders. The second form is partly driven by an external mechanism, in the sense that other teams use or become interested in the original group's expertise, and partly by the group's internal dynamism, in the sense that their articles, being published in more and more fields, have the potential to be applied in these other fields. In this contribution, we focus on basic counting measures as measures of diffusion. We introduce the notions of field diffusion breadth, defined as the number of for Essential Science Indicators (ESI) fields in which a set of articles is cited, and field diffusion intensity, defined as the number of citing articles in one particular ESI field. Combined effects of publications and citations can be measured by the Gini evenness measure. Our approach is illustrated by a study of mathematics at Tongji University (Shanghai, China).

### Introduction

Diffusion is one of the basic notions in publication and citation analysis and, as such, reviewed by Bar-Ilan (2006) in a recent review of the field of informetrics. A few years ago Rowlands (2002) proposed a new way to measure the diffusion of scientific knowledge. He focused on journals

as units of diffusion and introduced an indicator to measure the breadth of research influence. Assuming that  $U$  is the number of different journals that cited the articles published in the original journal (during a fixed time window), and that  $Cit$  denotes the total number of citations received by articles in the original journal (during the same fixed time window), then the diffusion indicator introduced by Rowlands, denoted as RDI, is defined as

$$RDI = \frac{U}{Cit} \quad (1)$$

The notation RDI stands for Rowlands diffusion index, but the symbol  $R$  could also stand for relative. Indeed, RDI is a relative indicator as it measures citations in different journals relative to the total number of citations. It is quite normal that RDI decreases over time. Empirically, Rowlands' diffusion factor is negatively correlated with the total number of citations received (Rowlands, 2002). Intuitively, diffusion in an absolute sense can only increase. This observation led Franssen (2004) to the definition of another diffusion index. Denoting this index as FDI (for Franssen diffusion index) it is defined as

$$FDI = \frac{U}{Pub} \quad (2)$$

where  $Pub$  denotes the number of publications in the journal (or, more generally, set of articles) under study. We see that Franssen replaced  $Cit$  in the definition of the RDI by  $Pub$ . When  $Pub$  is fixed, the FDI cannot decrease. Franssen's dif-

RESUMO EMPÍRICO

## Doctors' Online Information Needs, Cognitive Search Strategies, and Judgments of Information Quality and Cognitive Authority: How Predictive Judgments Introduce Bias Into Cognitive Search Models

Benjamin Hughes and Jonathan Wareham  
Department of Information Systems, ESADE, 60-62 Av Pedra Blava, Barcelona, Spain 08036  
E-mail: benjamin.hughes@alumni.es; wareham@acm.org

Indra Joshi  
Southampton University Hospitals NHS Trust (SUHT), Tremona Road, Southampton, Hampshire, United Kingdom, SO16 6YD. E-mail: indrajoshi@hotmail.com

Literature examining information judgments and Internet search behaviors notes a number of major research gaps, including how users actually make these judgments outside of experiments or researcher-defined tasks, and how search behavior is impacted by a user's judgment of online information. Using the medical setting, where doctors face real consequences in applying the information found, we examine how information judgments employed by doctors to mitigate risk impact their cognitive search. Diaries encompassing 444 real clinical information search incidents, combined with semistructured interviews across 35 doctors, were analyzed via thematic analysis. Results show that doctors, though aware of the need for information quality and cognitive authority, rarely make evaluative judgments. This is explained by navigational bias in information searches and via predictive judgments that favor known sites where doctors perceive levels of information quality and cognitive authority. Doctors' mental models of the Internet sites and Web experience relevant to the task type enable these predictive judgments. These results suggest a model connecting online cognitive search and information judgment literatures. Moreover, this implies a need to understand cognitive search through longitudinal or learning-based views for repeated search tasks, and

adaptations to medical practitioner training and tools for online search.

### Introduction

Information search is a process by which a person seeks knowledge about a problem or situation, constituting a major activity by the Internet's millions of users (Browne, Pitts, & Wetherbe, 2007). The Web is now a primary source of information for many people, driving a critical need to understand how users search or employ search engines (Jansen & Spink, 2006). Extensive literature examines not only behavioral models detailing the different moves or tactics during Internet search but also decision making or strategies described as *cognitive search models* (Navarro-Prieto, Scaife, & Rogers, 1999; Thatcher, 2006, 2008). The latter examines the cognitive aspects of the moves users employ to optimize their search performance, exploring elements such as expertise differences or judgments on when to terminate the search (e.g., Thatcher, 2006, 2008; Colby, 2002; Jalliet, 2003; Brown et al., 2007). This notion of judgment introduces a second stream of literature, Internet information judgments, where authors note that the use of predictive information judgments impacts decision making in search, based on an anticipation of a page's value before viewing it (Rish, 2002; Griffiths & Brophy, 2005).

Cognitive search models rarely explore the impact of predictive judgments. Most studies are based on tasks defined by researchers in experimental settings that are difficult to generalize to professional contexts or real use (Thatcher

Received June 13, 2009; revised August 12, 2009; accepted September 3, 2009

Additional Supporting Information may be found in the online version of this article.

© 2009 ASSS&T • Published online 10 December 2009 in Wiley InterScience

RESUMO EMPÍRICO

## Effects of Granularity of Search Results on the Relevance Judgment Behavior of Engineers: Building Systems for Retrieval and Understanding of Context

Panos Balatsoukas and Peter Demian  
Department of Civil & Building Engineering, Loughborough University, Leicestershire, LE 11 3TU, United Kingdom. E-mail: (P.Balatsoukas, P.Demian)@lboro.ac.uk

Granularity is a novel concept for presenting information in search result interfaces of hierarchical query-driven information retrieval systems in a manner that can support understanding and exploration of the context of the retrieved information (e.g., by highlighting its position in the granular hierarchy and exposing its relationship with relatives in the hierarchy). Little research, however, has been conducted on the effects of granularity of search results on the relevance judgment behavior of engineers. Engineers are highly motivated information users who are particularly interested in understanding the context of the retrieved information. Therefore, it is hypothesized that the design of systems with careful regard for granularity would improve engineers' relevance judgment behavior. To test this hypothesis, a prototype system was developed and evaluated in terms of the time needed for users to find relevant information, the accuracy of their relevance judgment, and their subjective satisfaction. To evaluate the prototype, a user study was conducted where participants were asked to complete tasks, complete a satisfaction questionnaire, and be interviewed. The findings showed that participants performed better and were more satisfied when the prototype system presented only relevant information in context. Although this study presents some novel findings about the effects of granularity and context on user relevance judgment behavior, the results should be interpreted with caution. For example, participants in this research were recruited by convenience and performed a set of simulated tasks as opposed to real ones. However, suggestions for further research are presented.

### Introduction

To date, several studies in interactive information retrieval and human-computer interaction have investigated the effects of the presentation of search result interfaces on user relevance judgment behavior (e.g., Drott 2000; Jobo & Jose, 2008; Rele & Duchowski, 2005; Sebrects, Vasilakis,

Miller, Cugini, & Laskowski, 1999). Findings of these studies have suggested that many interface components, such as the amount of information presented in metadata surrogates, the use of thumbnails, the highlighting of query terms, and the structure of search result interfaces, can have an impact on the time needed for users to find relevant information as well as the accuracy of their relevance judgments and user satisfaction. However, these studies were focused on the role of the search result interface as a means for identifying, selecting, and obtaining access to relevant information, neglecting other functions that a search result interface could serve such as *understanding of the context* of the retrieved information.

Some information-seeking studies, particularly those focused specifically on the needs of engineers, have provided evidence that engineers were especially interested in understanding the context of the retrieved information to gain orientation and inspiration (Kraaijenbrink, 2007). This behavior can be attributed to the highly structured and hierarchical nature of the content with which engineers interact to find and reuse information. In the case of query-driven information retrieval (IR) systems that provide access to repositories of engineering content, context is defined by the hierarchical relationship between a piece of information and other information components within the hierarchy (e.g., components and subcomponents in a parts hierarchy of a designed product). Therefore, the search result interfaces of these systems should not only communicate relevance in terms of the degree of similarity between a query and the full-text document (or its surrogate) but also support understanding of that information, such as its context and organization within the product parts hierarchy (Demian & Frachter, 2006a).

Granularity is a novel concept for organizing and presenting information in search result interfaces of query-driven IR systems in a manner that supports understanding and exploration of context. Although granularity has its roots in relational database technology, the modern concept of

Received September 7, 2009; revised October 7, 2009; accepted October 7, 2009

© 2009 ASSS&T • Published online 10 December 2009 in Wiley InterScience

RESUMO EMPÍRICO

## A Multilevel View on Interpersonal Knowledge Transfer

Minhyung Kang

*Samsung Economic Research Institute, 31st Floor, Samsung Life Seocho Tower 1321-15, Seocho2-dong, Seocho-gu, Seoul, 137-955, Korea. E-mail: mhyung.kang@samsung.com*

Young-Gul Kim

*Business School, Korea Advanced Institute of Science & Technology, 207-43 Cheongryangri2-dong, Dongdaemoon-gu, Seoul, 130-722, Korea. E-mail: domino2@unitel.co.kr*

Knowledge transfer among employees is a critical enabler of organizational learning. In this article, the direct and moderating effects of the multilevel (i.e., dyadic and individual levels) antecedents of knowledge transfer are examined based on social network and knowledge management research. By analyzing the survey responses from eight R&D groups of five firms using hierarchical linear modeling, we find that structural equivalence significantly influences interpersonal knowledge transfer at the dyadic level, even when strength of ties is controlled. At the individual level, the knowledge recipient's motivational factors such as group identification and the perceived expertise of colleagues show significant effects on knowledge transfer. Finally, the effect of strength of ties at the dyadic level is more influential when the recipient's group identification is low.

Krackhardt, 1992; Reagans & McEvily, 2003; Sabherwal & Sabherwal, 2005; Uzzi, 1997). On the other hand, a similarity of positions of the knowledge source and the recipient within a social network (i.e., structural equivalence) also facilitates knowledge transfer because it builds up common background knowledge from the shared contacts (Burt, 1987). The cognitive similarity from the common background knowledge reduces the cost of knowledge transfer (Reagans & McEvily, 2003). However, knowledge management researchers have paid much less attention to structural equivalence than to strength of ties. Reagans and McEvily (2003) considered both strength of ties and structural equivalence in their empirical study, but the influence of structural equivalence disappeared when the effect of strength of ties was introduced; however, these two concepts are fundamentally different. For example,

RESUMO EMPÍRICO

## User-Producer Interaction in Web Site Development: Motives, Modes, and Misfits

Frank J. van Rijnsoever and Jan Faber

*Innovation Studies Utrecht, Copernicus Institute, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands. E-mail: {F.vanrijnsoever, J.faber}@geo.uu.nl*

Marnix L.J. Brinkman and Marijn A. van Weele

*Science and Innovation Management, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands. E-mail: {M.L.J.Brinkman, M.A.vanWeele}@students.uu.nl*

This paper studies the role and effects of user-producer interaction (UPI) in the production of Web sites. The results of two empirical studies, one from a producer perspective and one from a user perspective, are presented. It is concluded that using UPI in the development of informative Web sites for a large public is a risky process, because it is hard to identify and define the most targeted users and their particular needs. The implications of these results are discussed.

Past research has shown that usability is an important factor in the adoption process of information technologies (Davis, 1989; van Rijnsoever, van Hameren, Walraven, & Dijk, 2009; Venkatesh & Davis, 2000). In order to better match products to the needs of their users, producers can involve potential users in the development of an envisioned product (Biemans, 1991; Franke & Piller, 2004; Lüthje, 2004; Von Hippel, 1976, 2005). The assessment of user needs is a very important part of that user involvement (Holt, 1988).

RESUMO EMPÍRICO

## Factors Affecting Shapers of Organizational Wikis

Dave Yates

College of Information Studies, University of Maryland, 4117J Hornbake South, College Park, MD 20742. E-mail: dyates@umd.edu

Christian Wagner

Department of Information Systems, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong. E-mail: iscw@cityu.edu.hk

Ann Majchrzak

Marshall School of Business, University of Southern California, Los Angeles, CA 90089-1421. E-mail: majchrza@usc.edu

New Web 2.0 technologies such as wikis permit any organizational member of a virtual community of practice (CoP) to dynamically edit, integrate, and rewrite content (what we call *knowledge shaping*) as well as contribute personal knowledge. Previous research on factors that motivate contribution in virtual CoPs has focused exclusively on factors explaining why people contribute their personal knowledge, with no research focused on why people make the knowledge-shaping contributions (rewriting, integrating, and restructuring pages) which are possible with wikis. We hypothesize that factors that explain frequency of contribution will be different for those who shape from those who contribute only their personal knowledge. The results support our hypotheses. In addition, we find that shapers are *not* more likely to be managers or members of a community's core group who might typically serve in an administrator role, contrary to prior expectations. The implications of using Web 2.0 tools to encourage this shaping behavior are discussed.

Introduction

last few decades (cf. Cox, 2005), we refer to virtual CoPs as a collective of voluntary knowledge contributors, distributed across traditional organizational boundaries, which enables members to share insights, experiences, and practical knowledge (Davenport & Hall, 2002; Schwen & Hara, 2003). In these CoPs, organizational members can build knowledge resources by using wikis to share, reuse, adapt, mobilize, and aggregate their knowledge into new intellectual capital (Nahapiet & Ghoshal, 1998; Sivilia et al., 2008). However, actual community building in the traditional sociological sense is often limited (Cox, 2005), as is direct interpersonal communication, which makes reliance on knowledge-sharing tools more critical for members to successfully share knowledge and reuse others' knowledge.

Research has suggested that organizational wikis may support knowledge sharing and reuse among members of virtual CoPs (Oreg & Nov, 2008; Wagner, 2005). Typically, these communities are self-organizing, with individuals who voluntarily choose to participate when participation is open

RESUMO EMPÍRICO

## From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web

Alberto Pepe, Matthew Mayernik, and Christine L. Borgman  
Center for Embedded Networked Sensing and Department of Information Studies,  
University of California, Los Angeles, CA 90095. E-mail: apepe@ucla.edu

Herbert Van de Sompel  
Digital Library Research and Prototyping Team, Los Alamos National Laboratory,  
Los Alamos, NM 87545

In the process of scientific research, many information objects are generated, all of which may remain valuable indefinitely. However, artifacts such as instrument data and associated calibration information may have little value in isolation; their meaning is derived from their relationships to each other. Individual artifacts are best represented as components of a life cycle that is specific to a scientific research domain or project. Current cataloging practices do not describe objects at a sufficient level of granularity nor do they offer the globally persistent identifiers necessary to discover and manage scholarly products with World Wide Web standards. The Open Archives Initiative's Object Reuse and Exchange data model (ORE) meets these requirements. We demonstrate a conceptual implementation of ORE to represent the scientific life cycles of embedded networked sensor applications in seismology and environmental sciences. By establishing relationships between publications, data, and contextual research information, we illustrate how to obtain a richer and more realistic view of scientific practices. That view can facilitate new forms of scientific research and learning. Our analysis is framed by studies of scientific practices in a large, multidisciplinary, multi-university science and engineering research center, the Center for Embedded Networked Sensing.

Introduction

Over the last century, scientific communication has relied heavily on the dissemination of papers, journal articles, and monographs. Libraries had well-established cataloging and access mechanisms to support these products

when all were in print form. Librarians devoted most of their cataloging efforts to monographs and journals, leaving the description of individual journal articles and papers to abstracting and indexing services. Now, scientific artifacts originate in digital form and exist in a much wider array of genres. These include manuscripts, publications, data, laboratory and field notes, instrument calibrations, preprints, grant proposals, talks, slides, patent applications, theses, dissertations, and genres specific to individual disciplines. Neither scientists nor librarians are coping well with this deluge (Bell, Hey, & Szalay, 2009; Borgman, 2007; Hey & Hey, 2006; Hey & Trefethen, 2005).

The processes by which science is conducted also have remained remarkably stable, if only at the most general level. Kenneth Mees (1918), writing in *Nature* almost a century ago, identified three stages of scientific knowledge production:

first, the production of new knowledge by means of laboratory research; secondly, the publication of this knowledge in the form of papers and abstracts of papers; thirdly, the digestion of the new knowledge and its absorption into the general mass of information by critical comparison with other experiments on the same or similar subjects. (p. 355)

Subsequent studies of scholarly communication have affirmed this general sequence of research activities for the sciences, based in the laboratory or not, and for many other empirical disciplines (Lanour, 1987; Meadows, 1974, 1998).

How are we to "digest and absorb" the deluge of digital artifacts into today's "general mass of information"? "Critical comparison" is as essential today as in Mees's day. Given the sophisticated tools now available, comparison should become easier. However, a wider range of scholarly artifacts is now publicly available, and those artifacts often exist in multiple versions or multiple stages of development. New tools, ser-

Received June 14, 2009; revised September 16, 2009; accepted October 5, 2009

© 2009 ASIS&T • Published online 15 December 2009 in Wiley InterScience

RESUMO EMPÍRICO

## An Empirical Analysis of Engineers' Information Behaviors

Mark A. Robinson

Centre for Socio-Technical Systems Design, Centre for Organizational Strategy, Learning and Change, Leeds University Business School, Maurice Keyworth Building, University of Leeds, Leeds LS2 9JT, United Kingdom. E-mail: m.robinson@lubs.leeds.ac.uk

The importance of the acquisition and provision of information within knowledge work such as engineering is widely acknowledged. This article reports an extensive empirical study of such information behaviors in engineers, using a novel and effective work sampling method. Seventy-eight design engineers each carried a portable handheld computer (PDA) for 20 working days. Once every hour, they were prompted to enter data concerning the task they were currently performing, including the information behaviors in which they were engaging. The resultant data represent a comprehensive picture of engineers' information behaviors and the percentage of their working time for which each of these behaviors accounts (55.75% in total). Specific hypotheses concerning the time spent engaged in these behaviors were also tested. Accordingly, it was found that participants spent substantially more time receiving information they had not requested than information they had, and this pattern was also reflected when they provided others with information. Furthermore, although there was no difference found between the time participants spent searching for information from other people compared with nonhuman sources, in the former case they spent relatively less time locating the information source and information within that source, and relatively more time engaged in problem solving and decision making. The results are discussed in terms of their implications for theory and organizational practice.

research reported here seeks to provide an extensive and intensive empirical analysis of such *information behaviors* (Huotari & Wilson, 2001) in a sample of design engineers. Broadly speaking, this article has two objectives: (a) to provide accurate and detailed data concerning the different types of information behaviors in which these professionals engage and, furthermore, the percentage of their working time for which each of these behaviors accounts; and (b) to test specific detailed hypotheses, concerning such behaviors, derived from previous theoretical and empirical work within the field. Previous related research is reviewed next.

### Engineers' Information Behaviors

A large volume of previous research has explored the information behaviors of engineers, professionals who can be classified as *knowledge workers* (Janz, Colquitt, & Noe, 1997). The most extensive review of the research literature in this area examined 456 technical reports, papers, and books, published between 1954 and 1994, addressing the information seeking and communication of engineers and scientists (King et al., 1994). The importance of such behaviors to these technical roles was demonstrated by the finding that 40 to 66% of working time was spent processing, communicating, and disseminating information. More recent observational research has specifically examined the acquisition and provision

RESUMO EMPÍRICO

## Scatter Matters: Regularities and Implications for the Scatter of Healthcare Information on the Web

Suresh K. Bhavnani

Center for Computational Medicine and Bioinformatics, Medical School, University of Michigan, Ann Arbor, MI 48109-1092. E-mail: bhavnani@umich.edu

Frederick A. Peck

School of Education, University of Colorado, Boulder, CO 80309-0249. E-mail: fredpeck@mail.com

Despite the development of huge healthcare Web sites and powerful search engines, many searchers end their searches prematurely with incomplete information. Recent studies suggest that users often retrieve incomplete information because of the complex scatter of relevant facts about a topic across Web pages. However, little is understood about regularities underlying such information scatter. To probe regularities within the scatter of facts across Web pages, this article presents the results of two analyses: (a) a *cluster analysis* of Web pages that reveals the existence of three page clusters that vary in information density and (b) a *content analysis* that suggests the role each of the above-mentioned page clusters play in providing comprehensive information. These results provide implications for the design of Web sites, search tools, and training to help users find comprehensive information about a topic and for a hypothesis describing the underlying mechanisms causing the scatter. We conclude by briefly discussing how the analysis of information scatter, at the granularity of facts, complements existing theories of information-seeking behavior.

### Introduction

one such Web site. However, although users of search engines and domain portals can easily find information for questions that have *specific* answers (e.g., "What is a melanoma?"; Eisenbach & Kohler, 2002), they have difficulty in finding answers for questions requiring *comprehensive* information composed of many facts (e.g., "What are the risk and prevention factors for melanoma?"; Bhavnani et al., 2003; Bhavnani et al., 2005b).

Finding comprehensive information about a healthcare topic is critical because an increasing number of people use information from the Web for a wide range of tasks leading to real-world actions and outcomes. For example, an estimated half of all American adults have searched online for healthcare information to become informed, to prepare for appointments and surgery, and to share information (Fox & Fallows, 2003). Furthermore, healthcare professionals have often emphasized the need for patients to get a comprehensive understanding of their disease (from a consumer's perspective) to improve their judgment in making healthcare decisions and to encourage higher treatment compliance (e.g., Sturdee, 2000; McGlynn et al., 2003).

Who is it difficult to find comprehensive information about

RESUMO EMPÍRICO

## Revisiting the Syntactical and Structural Analysis of Library of Congress Subject Headings for the Digital Environment

Kwan Yi

School of Library and Information Science, University of Kentucky, 331 Little Library Building, Lexington, Kentucky 40506-0224. E-mail: kwan.yi@uky.edu

Lois Mai Chan

School of Library and Information Science, University of Kentucky, 337 Little Library Building, Lexington, Kentucky 40506-0224. E-mail: loischan@uky.edu

With the current information environment characterized by the proliferation of digital resources, including collaboratively created and shared resources, Library of Congress Subject Headings (LCSH) is facing the challenges of effective and efficient subject-based organization and retrieval of digital resources. To explore the feasibility of utilizing LCSH in a digital environment, we might need to revisit its basic characteristics. The objectives of our study were to analyze LCSH in both syntactical and relational structures, to discover the structural characteristics of LCSH, and to identify problems and issues for the feasibility of LCSH as an effective subject access tool. This study reports and discusses issues raised by the syntactical and hierarchical structures of LCSH that present challenges to its use in a networked environment. Given the results of this study, we recommend a number of provisional future directions for the development of LCSH towards further becoming a viable system for digital and networked resources.

### Introduction

Since its inception in the 19th century, Library of Congress

have utilized LCSH for subject access of digital information in an automated manner: an automatic assignment of LCSH to cataloging records for electronic resources (Frank & Paynter, 2005); an association of LCSH to books in the Google Book Search project to facilitate access to a group of works on the same subject (Riley, 2007); and a potential linking of social tags and LCSH with the goal of organizing networked resources through a combination of tags and LCSH (Yi & Chan, 2009). Regarding the question of where LCSH stands currently in becoming a viable system in networked environments, Chan (2000, p. 172) has stated, "LCSH can become a versatile system that is capable of functioning in heterogeneous environments and can serve as the united basis for supporting diversified uses while maintaining semantic interoperability among them."

Furthermore, in a networked environment, a vast number of heterogeneous information resources in distributed information repositories have been organized with different controlled vocabularies and different organization systems. The issue of interoperability of controlled vocabularies

RESUMO EMPÍRICO

## A Framework of Automatic Subject Term Assignment for Text Categorization: An Indexing Conception-Based Approach

EunKyung Chung

Ewha Womans University, Library and Information Science, 11-1 Seodaemun-Gu Daehyun-Dong, Seoul, Korea 120-750. E-mail: echung@ewha.ac.kr

Shawne Miksa

University of North Texas, College of Information, Department of Library and Information Sciences, 1155 Union Circle 311068, Denton, TX 76203. E-mail: Shawne.Miksa@unt.edu

Samantha K. Hastings

University of South Carolina, School of Library and Information Science, 1501 Greene Street, Columbia, SC 29208. E-mail: shastings@sc.edu

The purpose of this study is to examine whether the understandings of subject-indexing processes conducted by human indexers have a positive impact on the effectiveness of automatic subject term assignment through text categorization (TC). More specifically, human indexers' subject-indexing approaches, or conceptions, in conjunction with semantic sources were explored in the context of a typical scientific journal article dataset. Based on the premise that subject indexing approaches or conceptions with semantic sources are important for automatic subject term assignment through TC, this study proposed an indexing conception-based framework. For the purpose of this study, two research questions were explored: To what extent are semantic sources effective? To what extent are indexing conceptions effective? The experiments were conducted using a Support Vector Machine implementation in WEKA (L.H. Witten & E. Frank, 2000). Using F-measure, the experiment results showed that cited works, source title, and title were as effective as the full text while a keyword was found more effective than the full text. In addition, the findings showed that an indexing conception-based framework was more effective than the full text. The content-oriented and the document-oriented indexing approaches especially were found more effective than the full text. Among three indexing conception-based approaches, the content-oriented approach and the document-oriented approach were more effective than the domain-oriented approach. In other words, in the context of a typical scientific journal article dataset, the objective contents and authors' intentions were more desirable for automatic subject term assignment via TC

Received August 28, 2008; revised October 9, 2008; accepted October 12, 2009.

© 2010 ASIS&T • Published online 26 January 2010 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21272

than the possible users' needs. The findings of this study support that incorporation of human indexers' indexing approaches or conceptions in conjunction with semantic sources has a positive impact on the effectiveness of automatic subject term assignment.

### Introduction

Subject representation of information entities through the use of subject indexing has been a practice in information organization for centuries. Subject terms or headings serve as subject access points of value to information users when searching information retrieval systems. Traditionally, the facilitation of subject access to information has been achieved by human indexers' assignment of subject terms to documents utilizing appropriate controlled vocabularies or thesauri. However, due to the increasing volume of information and the perpetual need to organize and give access to information by subject, there have been numerous endeavors to automatically assign subject terms to the documents by using the full text of the document. One way to assign subject terms automatically is through the use of text categorization (TC) using supervised machine learning algorithms. However, as Cunningham, Witten, and Litten (1999) noted, the models and properties of TC have been approached without reasonably solid understandings of how human indexers approach subject indexing. More specifically, research in TC focuses on statistical and probabilistic foundations with respect to document representation, parameter optimizations, and algorithm developments to improve effectiveness rather than basing it on understandings of subject indexing as a conceptual framework. Consequently, there has been little

RESUMO EMPÍRICO

## The Impact of Awareness and Accessibility on Expertise Retrieval: A Multilevel Network Perspective

Y. Connie Yuan

Department of Communication & Department of Information Science, 308 Kennedy Hall, Cornell University, Ithaca, NY 14853. E-mail: yy239@cornell.edu

Inga Carboni

Mason School of Business, College of William & Mary, Tyler Hall, Room 322, Williamsburg, VA, 23187-8795. E-mail: inga.carboni@mason.wm.edu

Kate Ehrlich

IBM T.J. Watson Research, 1 Rogers Street, Cambridge MA 02142. E-mail: kate@us.ibm.com

Building on the major premises of transactive memory (TM) theory as well as the recent multilevel extension to the original theory, this study examined the influence of perceived social accessibility of expertise providers, technological accessibility, and awareness of expertise distribution on expertise retrieval. Using social network data collected from a large global sales team, the study found that all three variables had significant impact on expertise retrieval at both the dyadic and individual levels. Our study confirmed the conceptual and theoretical value of approaching TM from a multilevel network perspective.

However, the simple proposition that equates *awareness* of expertise with *access* to expertise has been challenged by recent field studies (Casciaro & Lobo, 2008; Faraj & Sproull, 2000). For instance, Faraj and Sproull (2000) demonstrated that team performance was a function of access to expertise rather than the mere presence of expertise in the team, indicating that awareness of expertise by itself may not adequately explain performance. Yuan, Fulk, Monge, and Contractor (in press) found that awareness of "who knows what" needs to be supported by strong relationships to obtain actual access to expertise. Existing knowledge management research also has found that employee communication and knowledge sharing

### RESUMO EMPÍRICO

## Improving Wikipedia's Credibility: References and Citations in a Sample of History Articles

Brendan Luyt and Daniel Tan

Wee Kim Wee School of Communication & Information, Nanyang Technological University, Singapore. E-mail: brendan@ntu.edu.sg; daniel.tandz@gmail.com

This study evaluates how well the authors of Wikipedia history articles adhere to the site's policy of assuring verifiability through citations. It does so by examining the references and citations of a subset of country histories. The findings paint a dismal picture. Not only are many claims not verified through citations, those that are suffer from the choice of references used. Many of these are from only a few US government Websites or news media and few are to academic journal material. Given these results, one response would be to declare Wikipedia unsuitable for serious reference work. But another option emerges when we jettison technological determinism and look at Wikipedia as a product of a wider social context. Key to this context is a world in which information is bottled up as commodities requiring payment for access. Equally important is the problematic assumption that texts are undifferentiated bearers of knowledge. Those involved in instructional programs can draw attention to the social nature of texts to counter these assumptions and by so doing create an awareness for a new generation of Wikipedians and Wikipedia users of the need to evaluate texts (and hence citations) in light of the social context of their production and use.

#### Introduction

Wikipedia from very early in its history has had as one of its core policies governing article construction the require-

menting practice? From an analysis of the references and citations found in a subset of Wikipedia's history articles, this study answers that it is not. Not only are citations not used to verify all claims, those references that are used are primarily from a few United States government Websites and online media news. The findings, in fact, are dismal. But rather than dismiss Wikipedia from further consideration, this article argues that we need to pay more attention to the social context that creates the conditions for these references to be generally accepted among the Wikipedian community and by extension the many users of Wikipedia: the general commodification of much useful information on the Internet and a view of texts as undifferentiated bearers of knowledge. Once these problems are recognized, certain approaches to information literacy can be used to cultivate a new generation of Wikipedians and Wikipedia users better attuned to the complexities of reference practice and hence capable of rectifying the current lack of credibility that the online encyclopedia has among many people.

#### Review of the Literature

The ability to easily edit Wikipedia articles and instantaneously see the results reflected on the site are key features

### RESUMO EMPÍRICO

## An Exploratory Study of Visual and Psychological Correlates of Preference for Onscreen Subpixel-Rendered Text

Randolph G. Bias  
*School of Information, University of Texas at Austin, Austin, TX 78701. E-mail: rbias@ischool.utexas.edu*

Kevin Larson  
*Advanced Reading Technologies, Microsoft Corporation, One Redmond Way, Redmond, WA 98052.  
E-mail: Kevlar@microsoft.com*

Sheng-Cheng Huang and Paul R. Aumer-Ryan  
*School of Information, University of Texas at Austin, Austin, TX 78701. E-mail: huangsc@mail.utexas.edu;  
paul@ischool.utexas.edu*

Chris Montesclaros  
*Austin Retina Associates, 801 W. 38th Street, Austin, TX 78705. E-mail: Montesclarochris@hotmail.com*

Font-rendering technologies play a critical role in presenting clear and aesthetic fonts to enhance the experience of reading from computer screens. This article presents three studies investigating visual and psychological correlates of people's preferences toward different onscreen text enhancements such as ClearType developed by Microsoft. Findings suggested that (a) people's acuity and hue sensitivity were two major factors that affect their preferences to ClearType's color filtering of subpixels on fonts, and (b) specific personality traits such as disagreeableness also could correlate with people's impressions of different onscreen text enhancements that were used. These empirical data would inform digital typographers and human-computer interaction scientists who aim to develop better systems of onscreen reading.

growing. Nearly everyone agrees that reading from computers is not as pleasant as reading from a beautiful leather-bound book or from a glossy, well-designed magazine. Research has shown, over two decades, that reading performance from computers has been inferior to that from hard copy (e.g., Gould & Grischkowsky, 1984). The goal of this project is to improve the experience of reading from computers, and the goal of our research is to understand how selected perceptual and personality variables may correlate with user preference for subpixel rendered text.

A lot of technology is involved in placing letters on a screen. In the 15th century, Gutenberg automated the process of book making by creating reusable punches that imposed letters on a page. Modern computer fonts use high-resolution outlines of letters that are scaled down to the

RESUMO EMPÍRICO

## A Review of Factors Influencing User Satisfaction in Information Retrieval

Azzah Al-Maskari and Mark Sanderson  
*Department of Information Studies, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK. E-mail: maskeri@gmail.com; msanderson@shef.ac.uk*

The authors investigate factors influencing user satisfaction in information retrieval. It is evident from this study that user satisfaction is a subjective variable, which can be influenced by several factors such as system effectiveness, user effectiveness, user effort, and user characteristics and expectations. Therefore, information retrieval evaluators should consider all these factors in obtaining user satisfaction and in using it as a criterion of system effectiveness. Previous studies have conflicting conclusions on the relationship between user satisfaction and system effectiveness; this study has substantiated these findings and supports using user satisfaction as a criterion of system effectiveness.

### Introduction

The main aim of an information retrieval (IR) system is

of system effectiveness, but is influenced by other factors, including system output, user expectation and attitude, perceived ease of use and usefulness, system type, and task difficulty.

Although there is much research calling for the importance of user satisfaction as a criterion of IR system evaluation (e.g., Gatian, 1994; Gluck, 1996; Huffman & Hochster, 2007; Spärck Jones, 1981), there is no comprehensive study investigating factors influencing user satisfaction. For example, several studies examined the relationship between user satisfaction and system effectiveness<sup>1</sup> (e.g., Gluck, 1996; Huffman & Hochster, 2007; Johnson, Griffiths, & Hartley, 2003; Sandore, 1990; Thomas & Hawking, 2006; Turpin & Hersh, 2001); however, these studies did not consider user experience and expectation of the IR system under evaluation. In this article, we discuss results that emerged from

RESUMO EMPÍRICO

## National Study of Information Seeking Behavior of Academic Researchers in the United States

**Xi Niu and Bradley M. Hemminger**

School of Information and Library Science, University of North Carolina at Chapel Hill, NC 27599-3360.  
E-mail: xniu@email.unc.edu; brhm@is.unc.edu

**Cory Lowm**

NC State Library, North Carolina State University, Raleigh, NC 27695-7111. E-mail: corylowm@gmail.com

**Stephanie Adams**

School of Information and Library Science, University of North Carolina at Chapel Hill, NC 27599-3360.  
E-mail: sa15@unc.edu

**Cecelia Brown**

School of Library and Information Studies, University of Oklahoma, Norman, OK 73019.  
E-mail: cbrown@ou.edu

**Allison Leavel and Marinda McLure**

Colorado State University Libraries, Fort Collins, CO 80523-1019.  
E-mail: jallison.leavel, marinda.mclure@colostate.edu

**Audrey Powers**

Tampa Library, University of South Florida, Tampa, FL 33620-5400. E-mail: apowers@lib.usf.edu

**Michelle R. Tennant**

Health Science Center Libraries and UF Genetics Institute, University of Florida, Gainesville, FL 32610-0206.  
E-mail: brtennant@ufl.edu

**Tara Cataldo**

Marston Science Library, University of Florida, Gainesville, FL 32611-7011. E-mail: tcataldo@ufl.edu

As new technologies and information delivery systems emerge, the way in which individuals search for information is changing. To understand different aspects of researchers' information-seeking behavior, this article surveyed 2,063 academic researchers in natural science, engineering, and medical sciences from five research universities in the United States. A Web-based, in-depth questionnaire was designed to quantify researchers' information searching, information use, and information storage behaviors. Descriptive statistics are reported.

Additionally, analysis of results is broken out by institutions to compare differences among universities. Significant findings are reported, and implications are discussed because of increased utilization of electronic methods for searching, sharing, and storing scholarly content, as well as for utilizing library services. Generally speaking, researchers in the five universities had similar information-seeking behavior, with small differences because of varying academic structures and myriad library services provided at the individual institutions.

### Introduction

The advent of personal computers and the Internet followed by the introduction of online electronic journals and databases at the beginning of the 1990s led to the development of online academic resources and the transformation

Received May 16, 2009; revised August 27, 2009; accepted December 18, 2009

© 2010 ASSISAT • Published online 24 February 2010 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21307

JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 61(5):869-896, 2010

RESUMO EMPÍRICO

## Text-Based Video Content Classification for Online Video-Sharing Sites

**Chunmeng Huang, Tianjun Fu, and Hsinchun Chen**

AI Lab, Department of Management Information Systems, University of Arizona, Tucson, AZ 85721.  
E-mail: {chunmeng, tufu}@email.arizona.edu; hchen@eller.arizona.edu

With the emergence of Web 2.0, sharing personal content, communicating ideas, and interacting with other online users in Web 2.0 communities have become daily routines for online users. User-generated data from Web 2.0 sites provide rich personal information (e.g., personal preferences and interests) and can be utilized to obtain insight about cyber communities and their social networks. Many studies have focused on leveraging user-generated information to analyze blogs and forums, but few studies have applied this approach to video-sharing Web sites. In this study, we propose a text-based framework for video content classification of online video-sharing Web sites. Different types of user-generated data (e.g., titles, descriptions, and comments) were used as sources for online videos, and three types of text features (lexical, syntactic, and content-specific features) were extracted. These feature-based classification techniques (C4.5, Naive Bayes, and Support Vector Machine) were used to classify videos. To evaluate the proposed framework, user-generated data from candidate videos, which were identified by searching user-given keywords on YouTube, were first collected. Then, a subset of the collected data was randomly selected and manually tagged by users as our experiment data. The experimental results showed that the proposed approach was able to classify online videos based on users' interests with accuracy rate up to 92.2%, and all three types of text features contributed to discriminating videos. Support Vector Machine outperformed C4.5 and Naive Bayes techniques in our experiments. In addition, our case study further demonstrated that accurate video-classification results are very useful for identifying implicit cyber communities on video-sharing Web sites.

### Introduction

With the emergence of Web 2.0, Web 2.0 sites such as forums, blogs, video-sharing Web sites, and wikis have become increasingly popular in the past few years. User behavior in Web 2.0 communities has changed from just

browsing Web pages to generating and spreading their own content and ideas. As noted by O'Reilly (2005), one of the main features of Web 2.0 is the "architecture of participation," which refers to online content that is generated by those who are motivated by their own personal interests. Numerous user-generated data provide valuable personal and up-to-date information such as user preferences, sentiments, and opinions, which previously could be obtained only through surveys and interviews.

To obtain insight from user-generated information, the ability to collect and analyze the considerable quantity of information becomes a challenge. Classification technologies provide promising methods to organize data according to different perspectives. Many studies have used classification technologies to analyze text-based data collected from blogs and forums and obtain insights. For example, Abrams, Chen, and Nandanakar (2008) applied sentiment analysis to improve opinion classification of Web forums in multiple languages. Zheng, Li, Chen, and Huang (2006) adopted writing-style features to identify online authorship.

Like blogs and forums, video-sharing Web sites are an important part of Web 2.0. For example, YouTube, the world's largest video-sharing Web site, receives more than 65,000 videos and 100-million video views every day. Video-classification techniques can be used to improve user experiences with video-sharing Web sites by identifying videos more closely related to users' personal interests and distinguishing them from the many irrelevant videos that are obtained by using keyword searches alone.

Another challenging issue for Web 2.0 sites is the issue of illegal content such as child pornography or threatening content such as sites exhorting violence and extremism. Among these, violent extremist content is considered to be among the most dangerous, especially after the attack of September 11th. The U.S. government invests many resources in detecting potential terrorism and protecting the United States from extremist violence. Chen, Thoms, and Fu (2008) found that extremists use Web 2.0 as an effective platform to share resources, promote their ideas, and

Received August 19, 2009; revised November 12, 2009; accepted November 13, 2009

© 2010 ASSISAT • Published online 29 January 2010 in Wiley InterScience

RESUMO EMPÍRICO

## A Flexible Content-Based Image Retrieval Model and a Customizable System for the Retrieval of Shapes

Gloria Bordogna and Marco Pagani

Consiglio Nazionale delle Ricerche, Istituto per la Dinamica dei Processi Ambientali, via Pasubio 5, 24044 Dalmine (BG) Italy. E-mail: {gloria.bordogna, marco.pagani}@idpa.cnr.it

The authors describe a flexible model and a system for content-based image retrieval of objects' shapes. Flexibility is intended as the possibility of customizing the system behavior to the user's needs and perceptions. This is achieved by allowing users to modify the retrieval function. The system implementing this model uses multiple representations to characterize some macroscopic characteristics of the objects shapes. Specifically, the shape indexes describe the global features of the object's contour (represented by the Fourier coefficients), the contour's irregularities (represented by the multifractal spectrum), and the presence of concavities and convexities (represented by the contour scale space distribution). During a query formulation, the user can specify both the preference for the macroscopic shape aspects that he or she considers meaningful for the retrieval, and the desired level of accuracy of the matching, which means that the visual query shape must be considered with a given tolerance in representing the desired shapes. The evaluation experiments showed that this system can be suited to different retrieval behaviors, and that, generally, the combination of the multiple shape representations increases both recall and precision with respect to the application of any single representation.

### Introduction

Although image retrieval has been an active research field

shapes, etc. (Aigrain, Zhang, & Petkovic, 1996; Chang & Hsu, 1992; Del Bimbo, 1999; Feng, Siu, & Zhang, 2003; Gudivada & Raghavan, 1995; Jain, 1997; Smeulders et al., 2000). This second approach introduced a new perspective in image retrieval, and many systems have been developed, such as MULTOS (Conti & Rabiti, 1990), BlobWorld (Carson, Belongie, Greenspan, & Malik, 1997), QBIC (Flickner et al., 1995), PhotoBook (Penland, Picard, & Sclaroff, 1996), VisualSEEK (Smith & Chang, 1997), MARS (Rui, Huang, Ortega, & Mehrotra, 1998), Netra (Ma & Manjunath, 1997), SIMPLicity (Yamada, Pickering, Jeannin, & Cieplinski, 2001), VIPER (Chaboz, 2001), QVE (Hirata & Kato, 1992) and many others. A survey on CBIR can be found in Idris and Panchanathan (1997), Rui, Huang, and Chang (1999), Long, Zhang, and Feng (2003). These systems differ in the visual aspects of image that are considered (color, shape, etc.), in the features used to represent each visual aspect, and in the function used to match features for assessing similarity to a query. Nevertheless, CBIR leaves many open issues when it is put into practice (Feng et al., 2003; Liu, Zhang, Lu, & Ma, 2007). Specifically, in CBIR systems design, the modeling of the correspondence between the system and the human similarity judgment is a main issue (Berretti, Del Bimbo, & Pala, 2000; Del Bimbo & Pala, 1997; Hirata & Kato, 1992;

RESUMO EMPÍRICO

## The Effects of Background Information and Social Interaction on Image Tagging

Judit Bar-Ilan, Maayan Zhitomirsky-Geffet, Yitzchak Miller, and Snunith Shoham  
Department of Information Science, Bar-Ilan University

In this article, we describe the results of an experiment designed to understand the effects of background information and social interaction on image tagging. The participants in the experiment were asked to tag 12 pre-selected images of Jewish cultural heritage. The users were partitioned into three groups: the first group saw only the images with no additional information whatsoever, the second group saw the images plus a short, descriptive title, and the third group saw the images, the titles, and the URL of the page in which the image appeared. In the first stage of the experiment, each user tagged the images without seeing the tags provided by the other users. In the second stage, the users saw the tags assigned by others and were encouraged to interact. Results show that after the social interaction phase, the tag sets converged and the popular tags became even more popular. Although in all cases the total number of assigned tags increased after the social interaction phase, the number of distinct tags decreased in most cases. When viewing the image only, in some cases the users were not able to correctly identify what they saw in some of the pictures, but they overcame the initial difficulties after interaction. We conclude from this experiment that social interaction may lead to convergence in tagging and that the "wisdom of the crowds" helps overcome the difficulties due to the lack of information.

### Introduction

The term Web 2.0 was coined in 2004 (O'Reilly, 2005); in September 2005, Tim O'Reilly had already found 9.5 million

content. Such objects can be bookmarked Web pages (e.g., at delicious.com), blog postings (e.g., at technorati.com), books (e.g., at librarything.com), museum objects (e.g., www.steve.museum), user-generated videos (e.g., at Youtube), or images (e.g., at Flickr.com). Virtually all such tagging systems currently on the Web comprise the following components: (a) resources, the objects for which tags are assigned; (b) users, those who assign the tags; and (c) tags, the free text words or phrases added (Marlow, Naaman, Boyd, & Davis, 2006). In a recent paper (Ding et al., 2009), the large-scale characteristics of several popular tagging systems were studied. These characteristics included tag distribution and tag clusters of Delicious, Flickr, and YouTube. Trant (2009a) provides an extensive review on social tagging and folksonomies.

In the current study, we explore whether user-produced tags provide an accurate and reliable source of information for further effective image indexing and retrieval. We also investigate how various collaboration scenarios influence the users' point of view and consequently their resulting tags.

In the past, image annotation was a privilege of art experts like museum curators and lexicographers. However, such approach usually reflects a rather narrow point of view on the picture and employs a limited and purely professional vocabulary, thus it does not prove to be very helpful for nonprofessional users who try to search for images. Typically nonexpert users tend to provide keywords that reflect their

RESUMO EMPÍRICO

## With a Little Help From My Friends: Self-Interested and Prosocial Behavior on MySpace Music

Judd Antin and Matthew Earp

School of Information, University of California, Berkeley, 102 South Hall, Berkeley, CA 94720.

E-mail: {jantin, matt}@ischool.berkeley.edu

In this article, we explore the dynamics of prosocial and self-interested behavior among musicians on MySpace Music. MySpace Music is an important platform for social interactions and at the same time provides musicians with the opportunity for significant profit. We argue that these forces can be in tension with each other, encouraging musicians to make strategic choices about using MySpace to promote their own or others' rewards. We look for evidence of self-interested and prosocial "friending" strategies in the social network created by Top Friends links. We find strong evidence that individual preferences for prosocial and self-interested behavior influence friending strategies. Furthermore, our data illustrate a robust relationship between increased prominence and increased attention to others' rewards. These results shed light on how musicians manage their interactions in complex online environments and extend research on social values by demonstrating consistent preferences for prosocial or self-interested behavior in a multifaceted online setting.

profiles, connect to fans and other musicians, and use the site as a platform for marketing and music distribution.

As a part of a musician's career development, MySpace Music provides an opportunity for significant monetary rewards. Musicians can profit by marketing themselves and by monetizing Web traffic through ad placement and digital downloads. Web traffic, then, is a valuable resource on MySpace Music.

In situations where a valuable resource is present, there can be tension between one's own rewards and the rewards of others (Dawes & Messick, 2000). Ideally, musicians on MySpace Music might make decisions to reward both themselves and others. Often, however, individuals must choose one type of reward over another. Decision making in these situations can be (but is not always) zero-sum: benefiting others comes at the expense of benefiting oneself, and vice versa.

For the musician interested in furthering his own career, self-interest would dictate that he make decisions to maxi-

RESUMO EMPÍRICO

## The Adoption of University Library Web Site Resources: A Multigroup Analysis

Yong-Mi Kim

School of Library and Information Studies, University of Oklahoma, Schusterman Center, 4502 East 41st St., Tulsa, Oklahoma 74135-2553. E-mail: yongmi@ou.edu

University libraries invest a massive amount of resources in digitizing information for the Web, yet there is growing concern that much of this information is being underutilized. The present study uses the technology acceptance model (TAM) to investigate university library website resources (ULWR) usage. We categorize users based on academic roles and then analyze them as subgroups in order to observe different adoption patterns across groups. A total of 299 usable responses was collected from four different universities and across three populations: undergraduate, master, and doctoral student/faculty groups. The findings show that different library users indeed access ULWR for different reasons, resulting in a need for tailored managerial efforts. Overall, the extended TAM explains undergraduate students' usage best; the explanatory power of the model is significantly lower for the doctoral student/faculty group. Some of the findings challenge results reported in TAM research in other fields. The unexpected findings may result from the application of the model to a different context. Detailed theoretical implications and managerial guidance are offered.

### Introduction

Information technology (IT) has changed the way libraries

influence users' appropriation of such resources (Jankowska, 2004)?

The technology acceptance model (TAM), adapted from the theory of reasoned action (TRA), is a well-supported theory for explaining users' technology adoption behaviors (e.g., Davis, 1989; Venkatesh & Morris, 2000; Karahanna, Agarwal, & Angst, 2006). TAM posits that usage is determined by intention to use (IU) a particular system, which, in turn, is determined by perceived usefulness (PU) and perceived ease of use (PEOU). Although there is evidence that this model can be useful in a university library context and offer insight into how digitalized information can best be utilized (e.g., Heinrichs, Lim, Lim, & Spangenberg, 2007; Hong et al., 2002; Lin, Chan, & Wei, 2006), there are some issues to consider. First, research findings are reported based on a single user group (e.g., either undergraduate students or master's students) or a combination of undergraduate and master's students, while doctoral students and faculty members, the heaviest users of online resources, are overlooked. Subsequently, the findings may not offer a holistic understanding of the utilization of university library Web site resources (ULWR). Second, it is difficult to understand how users, whose academic tasks are distinct, differ in the adoption of

RESUMO EMPÍRICO

## Query Polyrepresentation for Ranking Retrieval Systems Without Relevance Judgments

Miles Efron

Graduate School of Library and Information Science, University of Illinois, 501 E. Daniel St., Champaign, IL, 61820. E-mail: mefron@illinois.edu

Megan Winget

School of Information, University of Texas, 1 University Station D7000, Austin, TX 78712. E-mail: megan@ischool.utexas.edu

Ranking information retrieval (IR) systems with respect to their effectiveness is a crucial operation during IR evaluation, as well as during data fusion. This article offers a novel method of approaching the system-ranking problem, based on the widely studied idea of polyrepresentation. The principle of polyrepresentation suggests that a single information need can be represented by many query articulations—what we call *query aspects*. By skimming the top  $k$  (where  $k$  is small) documents retrieved by a single system for multiple query aspects, we collect a set of documents that are likely to be relevant to a given test topic. Labeling these skimmed documents as putatively relevant lets us build pseudorelevance judgments without undue human intervention. We report experiments where using these pseudorelevance judgments delivers a rank ordering of IR systems that correlates highly with rankings based on human relevance judgments.

### Introduction

represented in a variety of ways. In this article, we rely on multiple query representations—what we call *query aspects*—to build pseudorelevance judgments in support of system ranking.

This problem is important because the bottleneck in established system ranking methods (e.g., the Cranfield method) lies in obtaining relevance judgments, a list of which documents are relevant to each query. Relevance judgments are often called *qrels*. Although documents and queries are relatively easy to gather, creating relevance judgments requires significant effort and resources. Thus, recent years have seen increased interest in methods for evaluating IR systems without human-generated relevance judgments.

This article presents a novel method for creating relevance judgments without human assessments. The idea is very simple: Given a query  $q$  and a corpus of documents  $D$  we want to create a set of *qrels* for  $q$  and  $D$ . To do this, we propose the simple idea of query aspects.

Query aspects are variations—rearticulations, generaliza-

RESUMO EMPÍRICO

## High-Speed Rough Clustering for Very Large Document Collections

Kazuaki Kishida

School of Library and Information Science, Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345 JAPAN. E-mail: kishida@slis.keio.ac.jp

Document clustering is an important tool, but it is not yet widely used in practice probably because of its high computational complexity. This article explores techniques of high-speed rough clustering of documents, assuming that it is sometimes necessary to obtain a clustering result in a shorter time, although the result is just an approximate outline of document clusters. A promising approach for such clustering is to reduce the number of documents to be checked for generating cluster vectors in the leader-follower clustering algorithm. Based on this idea, the present article proposes a modified Crouch algorithm and incomplete single-pass leader-follower algorithm. Also, a two-stage grouping technique, in which the first stage attempts to decrease the number of documents to be processed in the second stage by applying a quick merging technique, is developed. An experiment using a part of the Reuters corpus RCV1 showed empirically that both the modified Crouch and the incomplete single-pass leader-follower algorithms achieve clustering results more efficiently than the original methods, and also improved the effectiveness of clustering results. On the other hand, the two-stage grouping technique did not reduce the processing time in this experiment.

### Introduction

A crucial difficulty that impedes the use of document clustering in practice is high-computational complexity in applying the algorithms to large data sets. For example, hierarchical clustering generally requires calculation of similarity between all pairs of documents, i.e., if  $N$  documents are included in the set, then  $O(N^2)$  scores of a similarity measure have to be computed and recorded in a storage device. Although the target database is not large in terms of current computer technology, e.g.,  $N = 100000$ , it is very difficult to process  $O(N^2)$  document pairs within a reasonable time. Note that  $N^2 = 10000000000$  in this example. Even if the clustering system adopts a *single-pass algorithm*, which has been developed for alleviating the computational difficulty in the IR field, this algorithm may also not be efficient enough for clustering a large set of documents. The algorithm usually tries to efficiently generate clusters by scanning the document set just once (or twice) to keep the complexity low where  $O(LN)$  comparisons of subject representation between  $L$  clusters and all  $N$  documents are needed in each scan. However, the number of comparisons sometimes drastically increases, i.e.,  $O(LN)$  approaches to  $O(N^2)$  as  $L$  becomes large.

Consequently, it takes a very long time until the final result

RESUMO EMPÍRICO

## Constructing Business Profiles Based on Keyword Patterns on Web Sites

Liwen Vaughan  
Faculty of Information and Media Studies, University of Western Ontario, London, Ontario, N6A 5B7 Canada.  
E-mail: lvaughan@uwo.ca

Juan Tang and Jian Du  
Institute of Scientific and Technical Information of Shanghai, 1555 Huai Hai Zhong Road, Shanghai, 200031, China. E-mail: (jtang, jdu)@libnet.sh.cn

The study examined the possibility of constructing business profiles (specifically, product profiles) based on keyword patterns on various types of Web sites, including a company's own Web site, blog sites, and Web sites that have particular keywords and also hyperlinks pointing to company Web sites. To test the proposed methods, we selected China's four major oil companies and two other companies that have related products. We collected three rounds of data over a 7-month period from these three Web sources and analyzed the numbers of retrieved pages to construct business profiles. The business profiles constructed were checked against business information collected from other sources such as company annual reports and company newsletters to determine the correctness of the profiles and thus the usefulness of the proposed methods. We found that we can construct fairly accurate profiles by examining the frequency distribution of product keywords on company Web sites. Analyzing the frequency distribution of blogs on various topics was very useful in following major business events and developments during particular time periods. We also conducted qualitative content analysis for a sample of 454 Web pages retrieved from the three sources. Findings from the content analysis confirmed the conclusions from the quantitative analysis.

### Background and Purpose of the Study

The increasing role that the World Wide Web plays in our society has prompted an entirely new branch of academic research that has focused on various Web phenomena. The most commonly studied object in information science is

Web hyperlinks (e.g., Björneborn & Ingwersen, 2004; Ingwersen, 1998). Recent studies have begun to analyze other Web objects such as Weblogs (Bar-Ilan, 2005; Smith, 2007). The type of Web sites studied have varied, ranging from academic sites (e.g., Aguilto, Granados Ortega, & Prieto, 2006; Vaughan & Thelwall, 2005) to social sites (e.g., Thelwall, 2008) and political sites (Bar-Ilan & Eicherman, 2005; H. Chen et al., 2008; Holmberg, 2009; Thelwall, 2007). Studies of business sites or studies that have analyzed Web-related business phenomena are relatively fewer. Still, there are plenty of such studies, as reviewed next, from which we drew experience and developed the methodology for the current study.

Liu, Ma, and Yu (2001) were among the first to study business Web sites. They developed a method of gathering business information by monitoring Web sites of business competitors. Later studies analyzed how to obtain business information from Web hyperlinks (e.g., Reid, 2005; Vaughan & Wu, 2004) and a combination of hyperlinks and texts (Chau, Shiu, Chan, & Chen, 2007; Vaughan & You, 2009). Web messages and Weblogs also have been studied. Das and Sisk (2005) analyzed messages posted to stock boards, mostly the Yahoo! message board, using social network analysis. They found that network analysis of financial communities provides a way to base portfolio strategies. On the other hand, Tumarkin and Whitley (2001) studied the relationship between Internet message board activity and abnormal stock returns and trading volume. They found no causal link between message board activity and stock returns and volume; however, this study was carried out between 1999 and 2000, and the Web has changed substantially since then. The significance of blogs for business communities was noted as early as 2005 (Kirkpatrick, Roth, &

Received December 1, 2009; revised January 15, 2010; accepted January 15, 2010  
© 2010 ASSA/T • Published online 10 March 2010 in Wiley InterScience

RESUMO EMPÍRICO

## Contextual Cocitation: Augmenting Cocitation Analysis and its Applications

Alison Callahan and Stephen Hockema  
Faculty of Information, University of Toronto, 140 St. George St., Toronto, Ontario M5S 3G6.  
E-mail: alison.callahan@gmail.com; steve.hockema@utoronto.ca

Gunther Eysenbach  
Centre for Global eHealth Innovation, Toronto General Hospital, R. Fraser Elliott Building,  
4th Floor, room #4S435, 190 Elizabeth Street, Toronto, Ontario M5G 2C4.  
E-mail: geysenba@uhnres.utoronto.ca

In this work, a novel method of cocitation analysis, coined "contextual cocitation analysis," is introduced and described in comparison to traditional methods of cocitation analysis. Equations for quantifying contextual cocitation strength are introduced and their implications explored using theoretical examples alongside the application of contextual cocitation to a series of BioMed Central publications and their cited resources. Based on this work, the implications of contextual cocitation for understanding the granularity of the relationships created between cited published research and methods for its analysis are discussed. Future applications and improvements of this work, including its extended application to the published research of multiple disciplines, are then presented with rationales for their inclusion.

of the current literature of cocitation analysis, citer motivation, and classes of citation. Future applications of contextual cocitation analysis include improving the literature selection process for research-motivated reading, studying the relationships between cited online resources, and, on a grander scale, rethinking the resolution of our existing methodological tools for exploring the relationships that exist among published works on a finer-grained basis.

### Related Work

*Cocitation Analysis and Bibliographic Coupling*

Author cocitation analysis and document cocitation anal

RESUMO EMPÍRICO

## Recognizing Contributions in Wikis: Authorship Categories, Algorithms, and Visualizations

Ofer Arazy

Department of Accounting and Management Information Systems (AMS) School of Business,  
University of Alberta, Edmonton, AB T6G 2E8 Canada. E-mail: [ofer.arazy@ualberta.ca](mailto:ofer.arazy@ualberta.ca)

Eleni Stroulia

Computing Science Department, University of Alberta, Edmonton, AB T6G 2E8 Canada.  
E-mail: [elstrou@ualberta.ca](mailto:elstrou@ualberta.ca)

Stan Ruecker

Humanities Computing Program, Department of English and Film Studies and Office of Interdisciplinary  
Studies, University of Alberta, Edmonton, AB T6G 2E5 Canada. E-mail: [sruecker@ualberta.ca](mailto:sruecker@ualberta.ca)

Cristina Arias

Humanities Computing Program, Office of Interdisciplinary Studies, University of Alberta, Edmonton,  
AB T6G 2E5 Canada. E-mail: [cga@ualberta.ca](mailto:cga@ualberta.ca)

Carlos Fiorentino

Department of Art and Design, University of Alberta, Edmonton, AB T6G 2C9 Canada.  
E-mail: [carlos.florentino@gmx.net](mailto:carlos.florentino@gmx.net)

Veselin Ganev and Timothy Yau

Computing Science Department, University of Alberta, Edmonton, AB T6G 2E8 Canada.  
E-mail: [vganev,thyau@ualberta](mailto:vganev,thyau@ualberta)

Wikis are designed to support collaborative editing, without focusing on individual contribution, such that it is not straightforward to determine who contributed to a specific page. However, as wikis are increasingly adopted in settings such as business, government, and education, where editors are largely driven by career goals, there is a perceived need to modify wikis so that each editor's contributions are clearly presented. In this paper we introduce an approach for assessing the contributions of wiki editors along several authorship categories, as well as a variety of information glyphs for visualizing this information. We report on three types of analysis: (a) assessing the accuracy of the algorithms, (b) estimating the understandability of the visualizations, and (c) exploring wiki editors' perceptions regarding the extent to which such an approach is likely to change their behavior. Our findings demonstrate that our proposed automated techniques can estimate fairly accurately the quantity

of editors' contributions across various authorship categories, and that the visualizations we introduced can clearly convey this information to users. Moreover, our user study suggests that such tools are likely to change wiki editors' behavior. We discuss both the potential benefits and risks associated with solutions for estimating and visualizing wiki contributions.

### Introduction

Wiki, derived from the Hawaiian word for *fast*, is a Web-based collaborative authoring application (Leif & Cunningham, 2001; Wagner, 2004). In some ways, wikis are similar to discussion forums and blogs; both technologies enable communities to share information on the Web. At the same time, there is a distinct difference between these technologies. In forums and blogs, users post distinct entries that are appended sequentially, according to their timestamp, to a single page. As a result, each individual post is directly associated with its contributor. In wikis, on the other hand, users can edit any part of the content of a wiki page as a result

Received December 10, 2009; revised January 25, 2010; accepted January 25, 2010

RESUMO EMPÍRICO

## Search Behavior of Media Professionals at an Audiovisual Archive: A Transaction Log Analysis

Bouke Huurnink

Intelligent Systems Lab Amsterdam, University of Amsterdam, Science Park 107, Amsterdam,  
the Netherlands

Laura Hollink

Computer Science Department, FEW, VU University Amsterdam, De Boelelaan 1081A, Amsterdam,  
the Netherlands

Wietske van den Heuvel

Research and Development Department, Netherlands Institute for Sound and Vision, Sumatraaan 45,  
Hilversum, the Netherlands

Maarten de Rijke

Intelligent Systems Lab Amsterdam, University of Amsterdam, Science Park 107, Amsterdam,  
the Netherlands

Finding audiovisual material for reuse in new programs is an important activity for news producers, documentary makers, and other media professionals. Such professionals are typically served by an audiovisual broadcast archive. We report on a study of the transaction logs of one such archive. The analysis includes an investigation of commercial orders made by the media professionals and a characterization of sessions, queries, and the content of terms recorded in the logs. One of our key findings is that there is a strong demand for short pieces of audiovisual material in the archive. In addition, while searchers are generally able to quickly navigate to a usable audiovisual broadcast, it takes them longer to place an order when purchasing a subsection of a broadcast than when purchasing an entire broadcast. Another key finding is that queries predominantly consist of (parts of) broadcast titles and of proper names. Our observations imply that it may be beneficial to increase support for fine-grained access to audiovisual material, for example, through manual segmentation or content-based analysis.

### Introduction

Documentary makers, journalists, news editors, and other media professionals routinely require previously recorded

audiovisual material for reuse in new productions. For example, a news editor might wish to reuse footage shot by overseas services for the evening news, or a documentary maker describing the history of the Christmas tradition might desire footage from Christmas broadcasts in the 1930s. To complete production, the media professional must locate audiovisual material that has been previously broadcast in another context. One of the sources for reusable broadcasts is the audiovisual archive, which specializes in the preservation and management of audiovisual material (Edmondson, 2004). Although audiovisual material was once primarily stored on analog carriers, in recent years, audiovisual archives have started making their content available in digital format and enabling online access (e.g., Oomen, Verwoyen, Timmermans, & Heijmans, 2009; Wright, 2007). In such a digital environment, the media professional can search for and purchase multimedia material without leaving the comfort of his or her own office. In addition, with audiovisual acquisition being done through a digital interface, the archive can record information about the media professional's information-seeking process. Despite the fact that an increasing amount of audiovisual programming is being digitally

RESUMO EMPÍRICO