

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO**

Max Cirino de Mattos

**PROPOSTA DE UMA BASE DE CITAÇÕES DA LITERATURA
CIENTÍFICA POR MEIO DA EXTRAÇÃO AUTOMÁTICA DE DADOS
DO SCIELO**

Belo Horizonte

2013

MAX CIRINO DE MATTOS

**PROPOSTA DE UMA BASE DE CITAÇÕES DA LITERATURA
CIENTÍFICA POR MEIO DA EXTRAÇÃO AUTOMÁTICA DE DADOS
DO SCIELO**

Projeto apresentado ao Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Doutor em Ciência da Informação.

Linha de Pesquisa: Organização e Uso da Informação

Orientadora: Profa. Dra. Beatriz Valadares Cendón

Belo Horizonte

2013

Mattos, Max Cirino.

M435d Proposta de uma base de citações da literatura científica por meio da extração automática de dados do SciELO / Max Cirino Mattos. – 2013.

172 f., enc. : il., tabs., grafs. ; 29 cm.

Orientadora : Profa. Dra. Beatriz Valadares Cendón.

Área de concentração: Produção, organização e utilização da informação.

Linha de pesquisa: Organização e uso da Informação.

Tese (doutorado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Bibliografia: f. 138-149.

1. Bibliometria – Teses. 2. Ciência – Estatística - Teses. 3. Indicadores de ciência – Teses. 4. SciELO – Teses. I. Cendón, Beatriz Valadares. II. Universidade Federal de Minas Gerais. Escola de Ciência da Informação. III. Título.

CDU: 002:311



UFMG

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

FOLHA DE APROVAÇÃO

"PROPOSTA DE UMA BASE DE CITAÇÕES DA LITERATURA CIENTÍFICA POR MEIO DA EXTRAÇÃO AUTOMÁTICA DE DADOS DO SCIELO"

Max Cirino de Mattos

Tese 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 "**doutor em Ciência da Informação**", linha de pesquisa "**Organização e Uso da Informação**".

Tese aprovada em: 01 de outubro de 2013.

Por:

Prof. Dra. Beatriz Valadares Cendón - ECI/UFMG (Orientadora)

Prof. Dra. Cícera Henrique da Silva - FIOCRUZ/RJ

Prof. Dra. Maria Cristina Soares Guimarães - FIOCRUZ

Prof. Dr. Ricardo Hiroshi Caldeira Takahashi - ICEX/UFMG

Prof. Dra. Marlene Oliveira Teixeira de Melo - ECI/UFMG

Prof. Dra. Renata Maria Abrantes Baracho Porto - ECI/UFMG

Aprovada pelo Colegiado do PPGCI

Versão final Aprovada por

Prof. Dra. Renata Maria Abrantes Baracho Porto
Coordenadora

Prof. Dra. Beatriz Valadares Cendón
Orientadora



UFMG

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

ATA DA DEFESA DE TESE DE **MAX CIRINO DE MATTOS**, matrícula: 2010654956

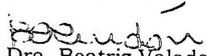
Às 9:00 horas do dia 01 de outubro de 2013, reuniu-se na Escola de Ciência da Informação da UFMG a Comissão Examinadora aprovada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação em 20/09/2013, para julgar, em exame final, o trabalho intitulado **Proposta de uma base de citações da literatura científica por meio da extração automática de dados do SciELO**, requisito final para obtenção do Grau de DOUTOR em CIÊNCIA DA INFORMAÇÃO, área de concentração: Produção, Organização e Utilização da Informação, Linha de Pesquisa: Organização e Uso da Informação. Abrindo a sessão, a Presidente da Comissão, Profa. Dra. Beatriz Valadares Cendón, 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:

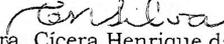
Profa. Dra. Beatriz Valadares Cendón - Orientadora	APROVADO
Profa. Dra. Cícera Henrique da Silva	APROVADO
Profa. Dra. Maria Cristina Soares Guimarães	APROVADO
Prof. Dr. Ricardo Hiroshi Caldeira Takahashi	APROVADO
Profa. Dra. Marlene Oliveira Teixeira de Melo	APROVADO
Profa. Dra. Renata Maria Abrantes Baracho Porto	APROVADO

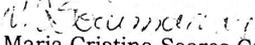
Pelas indicações, o candidato foi considerado APROVADO.

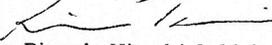
O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 01 de outubro de 2013.


Profa. Dra. Beatriz Valadares Cendón
ECI/UFMG


Profa. Dra. Cícera Henrique da Silva
FIOCRUZ/RJ

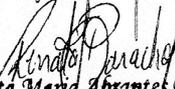

Profa. Dra. Maria Cristina Soares Guimarães
FIOCRUZ


Prof. Dr. Ricardo Hiroshi Caldeira Takahashi
ICEX/UFMG


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


Profa. Dra. Renata Maria Abrantes Baracho Porto
ECI/UFMG

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


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

Dedico este trabalho a Sibelle, Natália e Lucas, que tiveram a paciência e compreensão para que eu pudesse passar horas e horas no computador “para desenvolver a tese”.

AGRADECIMENTOS

À Sibelle, pela paciência, carinho e compreensão ao longo do desenvolvimento deste trabalho;

Aos meus pais, Angela e Roberto, pelo apoio incondicional;

À profa. Beatriz Valadares Cendón, pelos ensinamentos e pela tranquilidade e seriedade com que me ajudou no desenvolvimento desta tese e dos trabalhos publicados ao longo do curso;

Aos professores da ECI, que colaboraram para a construção deste trabalho, em especial Cátia Rodrigues Barbosa e Renata Maria Abrantes Baracho Porto;

Aos colegas da pós-graduação, que possibilitaram a ampliação de minha visão e conceitos (intelectuais e cotidianos) durante todo esse construtivo momento de troca de conhecimentos, em especial Wesley, Fátima, Edson e Rafael;

À Benildes Coura M. S. Maculan, pela excelente ajuda do “Manual de Normalização para o NITEG e o PPGCI da ECI-UFMG”;

À Gisele, Célio e Nely, pelo apoio e atenção dispensados durante todo o curso;

À Comarella e Mattos Consultores Associados, pela permissão de uso dos sistemas que viabilizaram a automação do protótipo;

À CAPES pelo auxílio financeiro representado pela bolsa REUNI.

“Nosso fascínio pela tecnologia nos fez esquecer o objetivo principal da informação: informar. Todos os computadores do mundo de nada servirão se seus usuários não estiverem interessados na informação que esses computadores podem gerar.”

DAVENPORT, 1998

RESUMO

Diversos autores ressaltam a importância da criação de uma base de citações – nos moldes do *Science Citation Index* (SCI) – para a visibilidade da produção científica local de países em desenvolvimento, inclusive para a produção de políticas científicas nacionais. Nesse sentido, a obtenção automática dos metadados dos artigos e referências citadas disponíveis no formato *eXtensible Markup Language* (XML) para a criação de uma base de citações – considerando como fonte primária a *Scientific Electronic Library On-line* (SciELO) – representa um passo inicial importante para a criação de um índice de citações para a América Latina e Caribe. A metodologia proposta trata da geração automática dessa base de citações, e o presente trabalho analisa os resultados encontrados nas 4 etapas iniciais dessa metodologia – a identificação dos periódicos; a obtenção dos dados estatísticos anuais (dados fonte) de cada periódico; a vinculação da(s) área(s) de conhecimento a cada periódico; e a criação do Módulo “Dados Cadastrais” do banco de dados – e nas três etapas finais: a identificação e armazenamento dos arquivos XML disponíveis no SciELO; a interpretação desses arquivos para extração dos metadados e informações sobre cada referência citada; e o armazenamento de todas as informações extraídas de cada arquivo XML no banco de dados (no Módulo “Base de Citações”). O teste inicial do protótipo criado foi realizado com o periódico *Perspectivas em Ciência da Informação* (PCI), apresentando a análise de 24 fascículos, 300 artigos, 7.714 citações, 579 resumos, 587 títulos, 2.358 palavras-chave, 686 autores de artigos e 10.394 autores identificados nas citações. A validação do protótipo foi realizada com a Coleção Saúde Pública, resultando em 14 periódicos, 14 editores, 1.335 fascículos, 23.780 artigos, 491.739 citações, 37.124 resumos, 44.696 títulos, 149.874 palavras-chave, 73.859 autores de artigos e 1.240.734 autores identificados nas citações. Nenhum procedimento de desambiguação de nomes de fontes ou autores foi realizado. As diferenças encontradas entre os números disponibilizados pelo SciELO nos dados fonte de cada periódico e os números levantados a partir da interpretação dos arquivos XML são explicadas, e algumas soluções propostas. O elevado índice de acerto na identificação das citações a partir dos arquivos XML comprovou a eficácia do processo desenvolvido. Entre os problemas identificados, destaca-se a diferença entre os dados fonte de um mesmo ISSN em coleções diferentes. Mais detalhes sobre a forma como o SciELO calcula os dados sobre número de fascículos, artigos e citações precisam ser investigados para a análise das diferenças encontradas. Pretende-se disponibilizar a base de citações gerada para a PCI em seu site. Nova pesquisa está em andamento, buscando a obtenção de todos os arquivos XML das coleções citadas para a construção de uma base de citações para cobertura da América Latina, Caribe e outras coleções do SciELO.

Palavras-chave: SciELO; base de citações; indicadores científicos nacionais.

ABSTRACT

Several authors emphasize the importance of creating a citation index - such as the Science Citation Index (SCI) – as an instrument for the production of national science policies and therefore for the promotion of local scientific development in less developed countries. The automatic retrieval of metadata of articles and references cited available in eXtensible Markup Language (XML) files to create this kind of index - using Scientific Electronic Library Online (SciELO) as a primary source - represents an important initial step for creating a Web of Science for Latin America and the Caribbean. The methodology used is based upon the automatic generation of such citations, and this research analyzes the results found in the initial stages of this methodology - identification of journals; obtaining the annual statistical data (source data) for each journal, the identification of areas of knowledge for each journal and the creation of the database module "Registration Data" - and the three final stages: identification and storage of XML files available in SciELO; interpreting these files for extracting metadata and information about each cited reference and the storage of all information from each XML file in the database module "Citation Index". The initial test of the prototype built was performed with the journal "Perspectives in Information Science" (PIS), presenting the analysis of 24 issues, 300 articles, 7,714 citations, 579 abstracts, 587 titles, 2,358 keywords, 686 authors of articles and 10,394 authors identified in citations. The validation of the prototype was performed with the Public Health Collection resulting in 14 journals, 14 publishers, 1,335 issues, 23,780 articles, 491,739 citations, 37,124 abstracts, 44,696 titles, 149,874 keywords, 73,859 authors of articles and 1,240,734 authors identified in citations. There were no disambiguation procedures for names of authors or sources. The differences between the values provided by the source data of SciELO for each journal and the numbers collected from the interpretation of the XML files are explained and some solutions are proposed. The high success rate in identifying metadata and citations from XML files proved the effectiveness of the prototype. Among the problems identified, one to highlight was the difference between the source data for the same ISSN in different collections. More details about how SciELO calculates the number of issues, articles and citations need to be investigated for the analysis of the differences found. It is intended to provide the citation index generated for PIS on its website. Another research study is underway which seeks to obtain all the XML files from listed collections of SciELO in order to construct a citation index for Latin America, the Caribbean and other collections of SciELO.

Key-words: SciELO; Citation Index; national scientific indicators.

LISTA DE FIGURAS

FIGURA 1 – Criação de uma base de citações a partir dos arquivos do SciELO.....	15
FIGURA 2 – Fases para a criação da base de citações do SciELO	18
FIGURA 3 – Processo automático de preparação dos dados cadastrais dos periódicos do SciELO	19
FIGURA 4 – Coleções do SciELO utilizadas para validação da primeira fase do protótipo	20
FIGURA 5 – Processo automático e contínuo de obtenção e interpretação dos arquivos XML	21
FIGURA 6 – Frequências de citações.....	32
FIGURA 7 – Estatísticas dos periódicos citados.....	33
FIGURA 8 – Estatísticas dos periódicos citantes	34
FIGURA 9 – Visão geral da metodologia: passos e modelagem de dados simplificada	49
FIGURA 10 – Relação de periódicos da Coleção Saúde Pública	51
FIGURA 11 – Lista de dados fonte: Revista Peruana de Medicina Experimental y Salud Pública	54
FIGURA 12 – Módulo “Dados Cadastrais”	56
FIGURA 13 – Resultado parcial da interpretação do XML SciELO: citações de um artigo	60
FIGURA 14 – Resumo da importação automática de arquivos XML: PCI 2012	61
FIGURA 15 – Tipos de citações encontradas: amostra.....	61
FIGURA 16 – Possível melhoria para a geração do arquivo XML: identificação de teses.....	62
FIGURA 17 – Possível erro na geração do arquivo XML: datas textuais	62
FIGURA 18 – Módulo “Base de Citações”	63
FIGURA 19 – Módulo “Dados Cadastrais” em números.....	65
FIGURA 20 – Periódicos SciELO agrupados por área de conhecimento.....	74
FIGURA 21 – Lista de dados fonte: Perspectivas em Ciência da Informação	76
FIGURA 22 – Resumo da importação de dados do SciELO: dados fonte e arquivos XML.....	78
FIGURA 23 – Artigos listados do ISSN 1413-9936 para o ano de 2013	79
FIGURA 24 – Referências citadas em um artigo de 2013 do ISSN 1413-9936	80
FIGURA 25 – Estrutura das <i>tags</i> XML dos arquivos interpretados	81
FIGURA 26 – Arquivos XML incorporados sem citações associadas	81
FIGURA 27 – Relação de artigos sem citações associadas, na PCI, para o ano de 2013	82
FIGURA 28 – Quantidade de registros incorporados no banco de dados: PCI.....	83
FIGURA 29 – Frequências de citações: PCI.....	86
FIGURA 30 – Estatística dos periódicos citados: PCI	87
FIGURA 31 – Estatística dos periódicos citantes: PCI	88
FIGURA 32 – Autores mais citados: PCI	90
FIGURA 33 – Palavras-chave mais utilizadas: PCI	91
FIGURA 34 – Autores que mais produziram: PCI	92
FIGURA 35 – Resumo dos dados de importação das citações da Coleção de Saúde Pública do SciELO	95
FIGURA 36 – Estrutura das <i>tags</i> XML dos arquivos interpretados: ISSN 0124-0064	96

FIGURA 37 – Resumo da importação de dados do SciELO: dados fonte X arquivos XML: ISSN 0124-0064	97
FIGURA 38 – Falha na estrutura das tags XML dos arquivos interpretados: ISSN 0124-0064	98
FIGURA 39 – Erro de acesso a arquivos XML: ISSN 0036-3634	99
FIGURA 40 – Resumo da importação de dados do SciELO: dados fonte X arquivos XML: ISSN 0036-3634	100
FIGURA 41 – Quantidade de registros incorporados no banco de dados: Coleção Saúde Pública	102
FIGURA 42 – Frequências de citações: Coleção Saúde Pública.....	103
FIGURA 43 – Estatística dos periódicos citados: Coleção Saúde Pública – visão parcial	104
FIGURA 44 – Estatística dos periódicos citantes: Coleção Saúde Pública – visão parcial.....	105
FIGURA 45 – Autores mais citados: Coleção Saúde Pública	107
FIGURA 46 – Palavras-chave mais utilizadas: Coleção Saúde Pública	108
FIGURA 47 – Autores que mais produziram: Coleção Saúde Pública (visão parcial).....	109
FIGURA 48 – Dados comparativos da Coleção Saúde Pública	116
FIGURA 49 – Resumo da importação de dados do SciELO: dados fonte X arquivos XML em 31/03/2013	119
FIGURA 50 – Resumo da importação de dados do SciELO: dados fonte X arquivos XML em 05/04/2013	120
FIGURA 51 – Resumo da importação de dados do SciELO para o ISSN 0036-3634	122
FIGURA 52 – Artigos incorporados na BC: ISSN 0034-8910 no período entre 31/03 e 05/04/2013	124
FIGURA 52 – Artigos incorporados na BC: ISSN 0034-8910 no período entre 31/03 e 05/04/2013 (cont.)	125
FIGURA 53 – Consulta artigos produzidos pelo autor BOCCATO na PCI	127
FIGURA 54 – Consulta pelo autor BOCCATO nas citações da PCI	128
FIGURA 54 – Consulta pelo autor BOCCATO nas citações da PCI (cont.)	129
FIGURA 54 – Consulta pelo autor BOCCATO nas citações da PCI (cont.)	130
FIGURA 55 – Lista de dados fonte na Coleção Saúde Pública: ISSN 0036-3634	150
FIGURA 56 – Lista de dados fonte na Coleção México: ISSN 0036-3634.....	150
FIGURA 57 – Lista de dados fonte na Coleção Saúde Pública: ISSN 0124-0064	151
FIGURA 58 – Lista de dados fonte na Coleção Colômbia: ISSN 0124-0064	151
FIGURA 59 – Lista de dados fonte na Coleção Saúde Pública: ISSN 0213-9111	152
FIGURA 60 – Lista de dados fonte na Coleção Espanha: ISSN 0213-9111	152
FIGURA 61 – Lista de dados fonte na Coleção Saúde Pública: ISSN 0864-3466	153
FIGURA 62 – Lista de dados fonte na Coleção de Cuba: ISSN 0864-3466	153
FIGURA 63 – Lista de dados fonte na Coleção Saúde Pública: ISSN 1135-5727	154
FIGURA 64 – Lista de dados fonte na Coleção Espanha: ISSN 1135-5727	154
FIGURA 65 – Lista de dados fonte na Coleção Saúde Pública: ISSN 1413-8123	155
FIGURA 66 – Lista de dados fonte na Coleção Brasil: ISSN 1413-8123	155
FIGURA 67 – Lista de dados fonte na Coleção Saúde Pública: ISSN 0102-311X	156
FIGURA 68 – Lista de dados fonte na Coleção Brasil: ISSN 0102-311X	156
FIGURA 69 – Lista de dados fonte na Coleção Saúde Pública: ISSN 0034-8910	157
FIGURA 70 – Lista de dados fonte na Coleção Brasil: ISSN 0034-8910	158

FIGURA 71 – Lista de dados fonte na Coleção Saúde Pública: ISSN 1415-790x	159
FIGURA 72 – Lista de dados fonte na Coleção Brasil: ISSN 1415-790x.....	159
FIGURA 73 – Dados comparativos da Coleção Saúde Pública: ISSN 1851-8265.....	160
FIGURA 74 – Dados comparativos da Coleção Saúde Pública: ISSN 1726-4634.....	161
FIGURA 75 – Dados comparativos da Coleção Saúde Pública: ISSN 1555-7960.....	161
FIGURA 76 – Dados comparativos da Coleção Saúde Pública: ISSN 1415-790x	162
FIGURA 77 – Dados comparativos da Coleção Saúde Pública: ISSN 1413-8123.....	163
FIGURA 78 – Dados comparativos da Coleção Saúde Pública: ISSN 1135-5727.....	164
FIGURA 79 – Dados comparativos da Coleção Saúde Pública: ISSN 1020-4989.....	165
FIGURA 80 – Dados comparativos da Coleção Saúde Pública: ISSN 0864-3466.....	166
FIGURA 81 – Dados comparativos da Coleção Saúde Pública: ISSN 0213-9111.....	167
FIGURA 82 – Dados comparativos da Coleção Saúde Pública: ISSN 0124-0064.....	168
FIGURA 83 – Dados comparativos da Coleção Saúde Pública: ISSN 0102-311X.....	169
FIGURA 84 – Dados comparativos da Coleção Saúde Pública: ISSN 0042-9686.....	170
FIGURA 85 – Dados comparativos da Coleção Saúde Pública: ISSN 0036-3634.....	171
FIGURA 86 – Dados comparativos da Coleção Saúde Pública	172

LISTA DE TABELAS E QUADROS

TABELA 1 – Periódicos encontrados em mais de uma Coleção do SciELO	66
TABELA 2 – Dados fontes do SciELO: diferenças para o ISSN 0036-3634.....	67
TABELA 3 – Dados fontes do SciELO: diferenças para o ISSN 0124-0064.....	68
TABELA 4 – Dados fontes do SciELO: diferenças para o ISSN 0213-9111.....	68
TABELA 5 – Dados fontes do SciELO: diferenças para o ISSN 0864-3466.....	69
TABELA 6 – Dados fontes do SciELO: diferenças para o ISSN 1135-5727.....	70
TABELA 7 – Dados fontes do SciELO: diferenças para o ISSN 1413-8123.....	70
TABELA 8 – Dados fontes do SciELO: diferenças para o ISSN 0102-311X	71
TABELA 9 – Dados fontes do SciELO: diferenças para o ISSN 0034-8910.....	72
TABELA 10 – Dados fontes do SciELO: diferenças para o ISSN 1415-790x.....	73
TABELA 11 – Resultados da importação XML: PCI 2006 a 2013	76
TABELA 12 – Resumo das diferenças entre o total de fascículos: dados fontes do SciELO	111
TABELA 13 – Resumo das diferenças entre o total de artigos: dados fontes do SciELO	112
TABELA 14 – Resumo das diferenças entre o total de citações concedidas: dados fontes do SciELO	113
TABELA 15 – Resumo das diferenças entre o total de citações recebidas: dados fontes do SciELO	114
TABELA 16 – Periódicos encontrados em mais de uma Coleção do SciELO e dados XML	115
QUADRO 1 – Informações sobre as Coleções do SciELO: domínios.....	53
QUADRO 2 – Informações sobre as Coleções do SciELO: prefixos	55
QUADRO 3 – Informações sobre as Coleções do SciELO	57
QUADRO 4 – Estrutura do arquivo XML do SciELO	59
QUADRO 5 – Periódicos da Coleção Saúde Pública do SciELO	93

LISTA DE ABREVIATURAS E SIGLAS

ABEC	Associação Brasileira de Editores Científicos
ABNT	Associação Brasileira de Normas Técnicas
BC	Base de Citações
BDTD	Biblioteca Digital de Teses e Dissertações
BIREME	Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CBBU	Comissão Brasileira de Bibliotecas Universitárias
CC	Ciência da Computação
CI	Ciência da Informação
CJP	<i>Citation Database for Japanese Papers</i>
CONICYT	<i>Comisión Nacional de Investigación Científica y Tecnológica</i>
CSP	Coleção Saúde Pública
ECI	Escola de Ciência da Informação
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FIOCRUZ	Fundação Oswaldo Cruz
GS	<i>Google Scholar</i>
GPL	<i>General Public License</i>
HTML	<i>HyperText Markup Language</i>
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IFES	Instituições Federais de Ensino Superior
ISI	<i>Institute of Scientific Information</i>
ISSN	<i>International Standard Serial Number</i>
LATTES	Plataforma Lattes
LILACS	Literatura Latino-Americana e do Caribe em Ciências da Saúde
MEC	Ministério da Educação
MEDLINE	<i>Medical Literature Analysis and Retrieval System Online</i>
MER	Modelo Entidade-Relacionamento
MTD-BR	Padrão brasileiro de metadados de teses e dissertações
OJS	<i>Open Journal Systems</i>
PAP	Programa de Aquisição Planificada de Periódicos para Bibliotecas Universitárias
PAAP	Programa de Apoio a Aquisição de Periódicos
PCI	Perspectivas em Ciência da Informação
PKP	<i>Public Knowledge Project</i>
PLOS	<i>Public Library of Science</i>
ProBE	Programa Biblioteca Eletrônica

SCI	<i>Science Citation Index</i>
SciELO	<i>Scientific Electronic Library Online</i>
SEER	Sistema Eletrônico de Editoração de Revistas
SGBD	Sistema de Gerenciamento de Banco de Dados
SICT	Sistemas de Informação em Ciência e Tecnologia
SIDA	Síndrome da Imunodeficiência Adquirida
UFMG	Universidade Federal de Minas Gerais
VINITI	<i>All-Union Institut for Science and Technical Information</i>
WOS	<i>Web of Science</i>
XML	<i>eXtensible Markup Language</i>

SUMÁRIO

1. INTRODUÇÃO	15
1.1 Justificativa	16
1.2 O protótipo	18
1.3 Problema e questão de pesquisa.....	23
1.4 Objetivos.....	23
1.5 Premissas e limites da pesquisa	23
1.6 Estrutura do trabalho	24
2. PERIÓDICOS CIENTÍFICOS, COLEÇÕES DIGITAIS E CIENTOMETRIA.....	25
2.1 Periódicos científicos e a internet.....	25
2.2 Bibliometria e Cientometria: criação e uso das bases de citações	29
2.2.1 <i>Web of Science</i>	31
2.2.2 Aplicações práticas	35
2.3 Bases de citações locais e o uso do SciELO	37
2.3.1 SciELO	38
2.4 Outras aplicações	41
2.4.1 Avaliação de coleções: o acervo do PORTAL CAPES.....	43
3. METODOLOGIA.....	48
3.1 Definição das ferramentas para desenvolvimento.....	50
3.2 FASE I: Obtenção dos dados do SciELO.....	50
3.2.1 Identificação dos periódicos de uma Coleção do SciELO	51
3.2.2 Obtenção dos dados estatísticos anuais de cada periódico.....	54
3.2.3 Vinculação da(s) área(s) de conhecimento de cada periódico.....	55
3.2.4 Criação do Módulo “Dados Cadastrais” do banco de dados	56
3.3 FASE II: Criação da Base de Citações.....	58
3.3.1 Levantamento dos arquivos XML de cada periódico	58
3.3.2 Interpretação dos arquivos XML do SciELO.....	59
3.3.3 Criação do Módulo “Base de Citações” do banco de dados	63
3.4 A seleção da amostra	64
4 RESULTADOS DA FASE I: COLEÇÕES DO SCIELO EM NÚMEROS.....	65
4.1 Resumo dos dados estatísticos: o Módulo Dados Cadastrais em números	65
4.2 Identificação dos periódicos e obtenção dos dados estatísticos.....	66
4.3 Vinculação às áreas de conhecimento.....	73
5 RESULTADOS DA FASE II: BASE DE CITAÇÕES DA PERSPECTIVAS EM CIÊNCIA DA INFORMAÇÃO	75
5.1 Obtenção dos dados do SciELO e criação da Base de Citações	75
5.2 Descrição da amostra: a PCI em números.....	83
6 RESULTADOS DA FASE II: BASE DE CITAÇÕES DA COLEÇÃO SAÚDE PÚBLICA	93
6.1 Obtenção dos dados do SciELO e criação da Base de Citações	93
6.2 Descrição da amostra: a Coleção Saúde Pública em números.....	101

7. ANÁLISE DOS RESULTADOS	110
7.1 Análise da obtenção dos dados do SciELO	110
7.2 Análise da criação da base de citações	117
7.3 Falta de padronização de nomes de autores e fontes	126
8. CONCLUSÕES E RECOMENDAÇÕES	132
8.1 Próximos passos.....	133
8.2 Futuras pesquisas.....	133
8.2.1 Desambiguação automática de nomes de autores e fontes.....	134
8.2.2 Avaliação automática de coleções digitais: uma proposta para o PORTAL CAPES	136
8.3 Considerações finais: a relevância do trabalho realizado.....	136
REFERÊNCIAS	138
ANEXO A – CONSULTAS AOS DADOS FONTE DO SCIELO.....	150
ANEXO B – CONSULTAS AOS DADOS DA BASE DE CITAÇÕES	160

1. INTRODUÇÃO

A pertinência à Ciência da Informação (CI) dos estudos bibliométricos especificamente relacionados à produção científica pode ser identificada, entre outras, nas reflexões apontadas por Garfield (1970; 1972; 1979; 1992; 1995; 1996). Alvarado (2007) conclui – a partir de uma detalhada revisão de literatura – que existe o reconhecimento da bibliometria e que ela pode ser tratada como uma subárea autônoma da CI que apresenta como importantes componentes a Lei de Lotka, a Lei de Zipf e a Lei de Bradford, bem como as análises de citações.

Interessa ao desenvolvimento deste trabalho particularmente o tema base de citações (BC), que é um elemento essencial para a realização de estudos bibliométricos que disponibiliza informações estruturadas a serem usadas para diversos tipos de análise da coleção de periódicos a que se refere.

A base de citações de maior destaque atualmente é o *Science Citation Index* (SCI), criado pelo *Institute for Scientific Information* (ISI) – atualmente *Thomson Reuters*. Genericamente, o conjunto de bases de dados de referências compiladas pelo ISI é denominado *Web of Science* (WoS), nome cunhado para a versão desse conjunto de bases disponibilizada na internet. Sua descrição – principalmente relacionada ao processo de sua criação – será detalhada no Capítulo 2.

Ao longo deste trabalho, os termos SCI, ISI e WoS foram usados de acordo com a forma adotada pelos autores consultados, e traduzem no contexto desta pesquisa o mesmo significado – um conjunto de bases de citações.

Em função do caráter seletivo do conjunto de bases de referências criadas pelo ISI – o que gera limitações para o suporte a políticas científicas locais – o presente trabalho propõe uma metodologia para a criação de uma base de citações, de forma automática e contínua, a partir dos arquivos disponíveis na base *Scientific Electronic Library Online* (SciELO), de acordo com a FIG. 1 a seguir:

FIGURA 1 – Criação de uma base de citações a partir dos arquivos do SciELO



Fonte: Desenvolvida pelo autor

1.1 Justificativa

De acordo com informações divulgadas no *site* do Senado Federal, a análise de trabalhos divulgados por brasileiros trabalhando no Brasil em revistas de circulação internacional cadastradas no SCI mostra um salto de 20 mil trabalhos publicados em 2007 para 30.415 trabalhos em 2008. Entretanto, de acordo com o professor Carlos Cruz, da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), esse salto é resultado, em grande parte, do cadastramento pelo ISI de novas revistas editadas no Brasil, e não se deve necessariamente a um efetivo aumento da produção científica. Conforme o pesquisador, a razão de crescimento do número de publicações vem caindo nos últimos anos: entre 1994 e 1998 foi de 18% ao ano; entre 1998 e 2002 foi de 9,3%; entre 2003 e 2009, foi de apenas 6% ao ano. (BRASIL, 2012)

Medir as publicações científicas que não se encontram indexadas no SCI é um processo importante, entretanto Meneghini (1998) apresenta uma análise da produção científica nacional e ressalta a carência de bases de dados brasileiras em informação científica nos moldes do SCI. De acordo com o autor, apenas cerca de 20% da produção brasileira encontra-se referenciada neste índice, e em relação aos 80% restantes ele aponta no mínimo três problemas: a falta de informações sobre a qualidade dessa produção, sobre o impacto de sua circulação e a falta de visibilidade internacional. O autor aborda o Projeto SciELO como instrumento para tornar a produção científica nacional mais visível internacionalmente, e acessível via meio eletrônico, criando uma base de dados que permita a avaliação dessa produção científica e a definição de estratégias e de políticas de gestão científica no Brasil.

De acordo com Cendón, Guimarães, Silva, Oliveira, Mattos, Santana, e Fernandes (2012),

a produção de indicadores que possam medir e avaliar a produção científica brasileira passa necessariamente pela existência de um índice de citações, nos moldes daqueles produzidos pelo *Institute of Scientific Information* (ISI), que contenha as referências citadas pelos pesquisadores brasileiros em suas produções científicas. Tais indicadores podem ser balizadores de políticas científicas nacionais, entre várias outras aplicações.

Os autores ainda afirmam que o Brasil desconhece o impacto real da produção científica da grande maioria dos seus cientistas e outras métricas que permitam analisar a produção científica nacional, o que parece ser comum “particularmente em áreas onde por tradição, a produção científica é disseminada principalmente, ou mesmo totalmente, em veículos nacionais ou de outros países periféricos” (CENDÓN, GUIMARÃES, SILVA, OLIVEIRA, MATTOS, SANTANA, e FERNANDES, 2012).

Para Guimarães, Silva, Santana, Braga, Bochner e Goldbaum (2011, p. 5), a deficiência da cobertura dos índices existentes (a exemplo da *Web of Science*) pode ocasionar deformações “nos processos de gestão das atividades científicas em contexto local”, o que levou vários países a buscar o desenvolvimento de índices de citações locais: China (XIN-NING, 2001), Polônia (WEBSTER, 1998) e União Europeia (GOGOLIN *et al.*, 2003) são exemplos citados.

Uma das dificuldades para a construção de uma base de citações é o enorme volume de dados que precisa ser compilado – o que praticamente inviabiliza o processo manual, sendo necessária a sua automação. Nesse sentido, Garfield, já em 1972, explicava que desde 1927 diversos autores mapeavam partes da rede existente de periódicos científicos, mas não existia um mapa geral desses periódicos – e a dificuldade prática para compilar o grande volume de dados de forma manual era o grande desafio a ser vencido.

Em outras palavras, vencer esse desafio significa aplicar a tecnologia como ferramenta de processamento dessas informações – ou seja, usar a tecnologia como recurso essencial para o desenvolvimento da CI, e não necessariamente abordar o tema sob a ótica da Computação.

Diretamente focado na carência de bases de citações nacionais e na necessidade de sua geração automática, o principal objetivo desta investigação é a proposição de uma metodologia para a criação de uma base de citações automática para as coleções do SciELO.

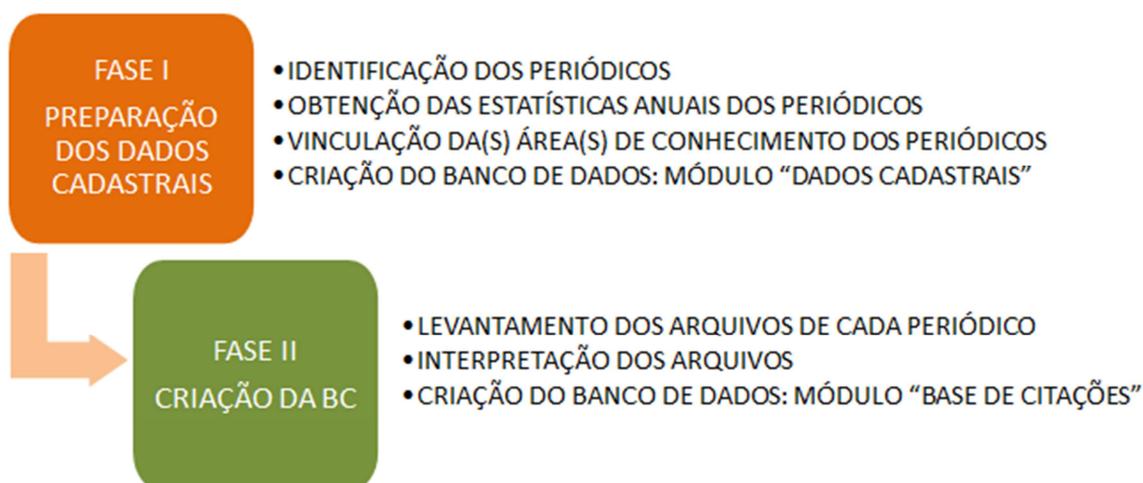
A escolha do SciELO deveu-se à sua cobertura nacional e internacional com foco em países de língua latina – a maior base de acesso livre mundial que contém coleções de periódicos dos seguintes países: Argentina, Brasil, Chile, Colômbia, Costa Rica, Cuba, Espanha, México, Portugal e Venezuela, além do processo iniciado na África do Sul, Bolívia, Paraguai, Peru e Uruguai¹. Além dessa cobertura, cabe ressaltar que a disponibilidade de acesso a arquivos no formato *eXtensible Markup Language* (XML) no SciELO foi essencial para a identificação automática dos metadados dos artigos e das referências usadas na sua produção.

¹ Disponível em: <<http://www.scielo.org/php/index.php>>. Acesso em: 01 mar. 2013.

1.2 O protótipo

Para a consecução do objetivo proposto foi elaborado um protótipo que trata da obtenção e interpretação do conteúdo dos arquivos XML disponíveis no SciELO. A metodologia proposta para a criação da base de citações do SciELO apresenta duas fases: (I) a obtenção de informações estatísticas anuais de cada periódico do SciELO para composição dos seus dados cadastrais, e (II) a obtenção e interpretação dos arquivos XML de cada um dos periódicos para composição da base de citações:

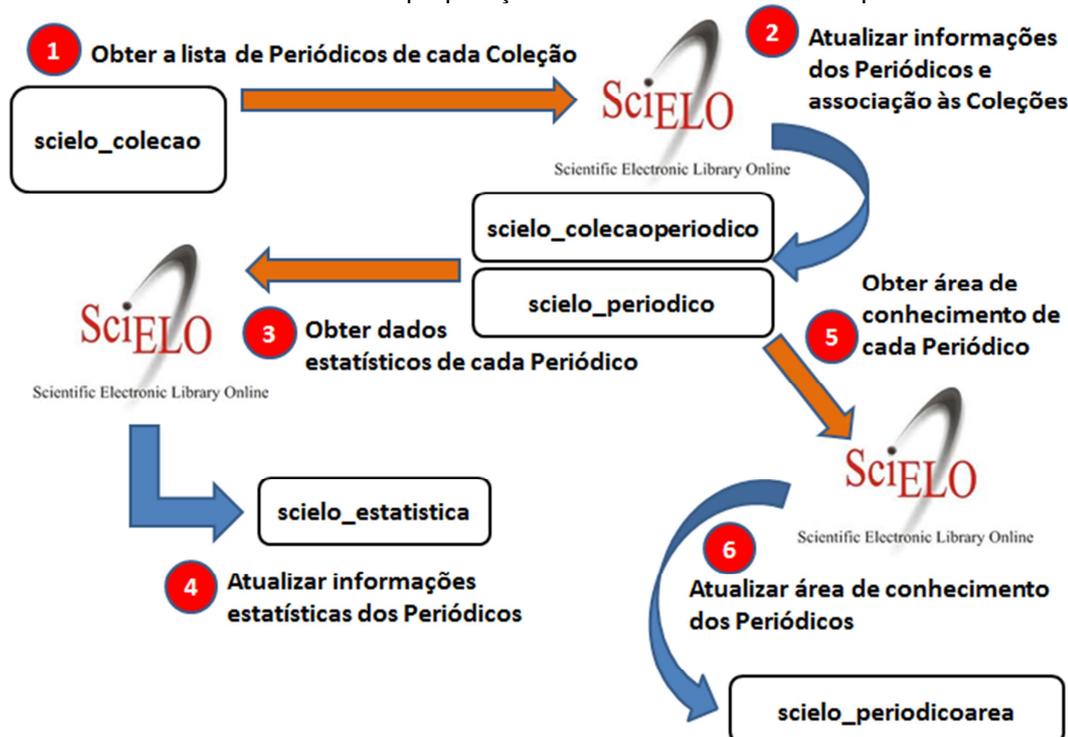
FIGURA 2 – Fases para a criação da base de citações do SciELO



Fonte: Desenvolvida pelo autor

A primeira fase está representada na FIG. 3 – após a obtenção automática da relação de periódicos de uma coleção, o sistema grava as informações estatísticas anuais disponíveis no SciELO; em seguida armazena a(s) área(s) de conhecimento à(s) qual(is) cada periódico pertence:

FIGURA 3 – Processo automático de preparação dos dados cadastrais dos periódicos do SciELO



Fonte: Desenvolvida pelo autor

Para o teste desta fase o *corpus* utilizado contemplou 10 coleções de periódicos do SciELO, com o total de 903 periódicos distintos conforme a FIG. 4 a seguir. A metodologia é detalhada no Capítulo 3, e os resultados encontrados para esta fase são apresentados no Capítulo 4.

FIGURA 4 – Coleções do SciELO utilizadas para validação da primeira fase do protótipo

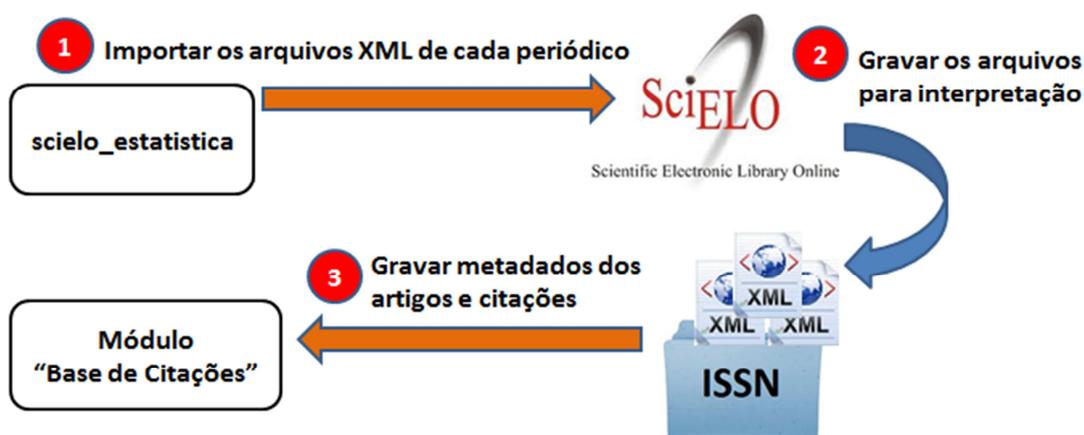
	Periódicos	Fascículos		Artigos		Citações Concedidas		Citações Recebidas		Correntes	Não correntes
10 coleções	903	26.984	100,00%	337.651	100,00%	8.875.927	100,00%	717.348	100,00%	780	123
<input checked="" type="checkbox"/> Brasil	316	13.627	50,50%	199.435	59,07%	5.265.923	59,33%	497.932	69,41%	271	45
<input checked="" type="checkbox"/> Chile	102	2.781	10,31%	30.034	8,89%	781.506	8,80%	38.050	5,30%	91	11
<input checked="" type="checkbox"/> Colombia	152	2.263	8,39%	23.656	7,01%	718.746	8,10%	21.250	2,96%	152	0
<input checked="" type="checkbox"/> Cuba	46	1.594	5,91%	17.300	5,12%	305.682	3,44%	23.633	3,29%	42	4
<input checked="" type="checkbox"/> Espanha	51	1.770	6,56%	15.697	4,65%	431.554	4,86%	24.820	3,46%	33	18
<input checked="" type="checkbox"/> Mexico	110	1.378	5,11%	12.846	3,80%	410.261	4,62%	17.017	2,37%	103	7
<input checked="" type="checkbox"/> Portugal	43	632	2,34%	5.630	1,67%	54.230	0,61%	2.726	0,38%	25	18
<input checked="" type="checkbox"/> Venezuela	53	1.214	4,50%	11.056	3,27%	279.904	3,15%	12.577	1,75%	33	20
<input checked="" type="checkbox"/> Saude Publica	15	1.419	5,26%	18.435	5,46%	508.175	5,73%	77.741	10,84%	15	0
<input checked="" type="checkbox"/> Africa do Sul	24	306	1,13%	3.562	1,05%	119.946	1,35%	1.602	0,22%	24	0

Fonte: desenvolvida pelo autor²

² Disponível em: <http://cmca.srv.br/prototipo/metabuscaador_index.php>. Acesso em: 12 jun. 2013

Na segunda fase, a partir dos dados estatísticos gravados na fase anterior para cada periódico, são obtidos e interpretados os arquivos XML do SciELO. Essa etapa grava as informações detalhadas de cada arquivo XML na base de citações, conforme a FIG. 5:

FIGURA 5 – Processo automático e contínuo de obtenção e interpretação dos arquivos XML



Fonte: Desenvolvida pelo autor

A metodologia de obtenção e interpretação desses arquivos é detalhada no Capítulo 3. O *corpus* usado para teste da segunda fase compreendeu todos os arquivos disponíveis no SciELO do periódico *Perspectivas em Ciência da Informação* (PCI), única revista da área de CI indexada no SciELO à época do desenvolvimento. A segunda amostra, utilizada para a validação do protótipo, contemplou os periódicos de uma coleção do SciELO: *Coleção Saúde Pública* (CSP). Os resultados desses testes encontram-se, respectivamente, nos Capítulos 5 e 6.

Os passos dos processos descritos na FIG. 3 e na FIG. 5 garantem a atualização automática e contínua das informações necessárias à composição da base de citações – ou seja, qualquer novo periódico disponibilizado pelo SciELO ou novos arquivos de determinado fascículo disponibilizados são identificados, importados e interpretados automaticamente.

É importante ressaltar que, apesar da ênfase na produção de uma ferramenta tecnológica, a geração de uma base de citações automática para o SciELO representa um passo inicial importante para diversas iniciativas de pesquisa e desenvolvimento científico, entre elas a própria atuação dos pesquisadores enquanto usuários dessa informação para o levantamento de fontes e autores, e a disponibilização de subsídios para a gestão do conhecimento científico – temas importantes para a CI.

Assim, esta pesquisa contribui de forma importante para o desenvolvimento da CI não a partir da análise do conteúdo da base de citações para uma ou outra área, mas antes pela disponibilização de um volume de informações e cobertura de áreas de

conhecimento que permitirão o desenvolvimento de inúmeras pesquisas – essas sim, específicas sobre áreas, autores e fontes de informações. Sem essa base de citações, sem esse passo tecnológico, toda a análise dos periódicos continuaria fragmentada em função do desafio descrito por Garfield: a compilação manual dos dados.

A abordagem da Ciência da Computação (CC) – que não é o caso deste trabalho – para essas ferramentas geralmente trata de algoritmos, *performance* de sistemas, normalização de bases de dados e análises técnicas e matemáticas abundantes. Entretanto, essas ferramentas podem ser usadas em diversas áreas como suporte ao tratamento de grandes volumes de informações e suporte à expansão das capacidades de comunicação (MATTOS, 2012a).

No caso específico de integração da CC com a CI, o desenvolvimento dessas ferramentas apresenta-se como uma possibilidade concreta de avanço para a CI ao permitir o tratamento de grandes volumes de informações, seja na organização, classificação, indexação, recuperação, definição de ontologias ou análise semântica de textos, entre outras possibilidades. O próprio desenvolvimento da base SCI somente foi possível a partir do uso do armazenamento magnético (GARFIELD, 1972).

É dessa forma que se apresenta neste trabalho o desenvolvimento de uma ferramenta tecnológica: não do ponto de vista técnico ou matemático, mas antes como um instrumento, um produto concreto resultante da combinação de conhecimentos nas duas áreas. Essa combinação tem sido abordada ao longo do tempo por alguns autores, entre eles Lourenço (2005), que abordou o Modelo Entidade Relacionamento (MER) e o padrão brasileiro de metadados de teses e dissertações (MTD-BR); Bohmerwald (2005), que apresentou uma visão integrada entre a CI (estudos de usuários) e a CC (usabilidade) para tratar da “análise mais completa de um sistema desenvolvido para Web, se a pesquisa abordar características tanto do estudo de usuários (mais especificamente do comportamento de busca), quanto do estudo da usabilidade.” (p. 95); e Maia (2013), que tratou da interconexão entre o MER e o Sistema Categorial de Ranganathan para a modelagem conceitual.

Nesse sentido, a ferramenta tecnológica aqui descrita apresenta-se como um importante produto de informação – de forma similar ao início do desenvolvimento do SCI – que poderá fomentar o desenvolvimento de uma base de citações para a América Latina e Caribe (MATTOS e CENDÓN, 2013) – uma aplicação prática importante do aspecto “Social Aplicado” da CI.

1.3 Problema e questão de pesquisa

Devido (I) à carência de bases de citações nos moldes do SCI que ofereçam dados de periódicos não contemplados pelo ISI e (II) à necessidade de automação do processo devido ao enorme volume de dados a ser tratado, e considerando que o SciELO contém vários periódicos não indexados no ISI e pertencentes a diversas áreas de conhecimento e países, o presente trabalho coloca o seguinte problema de pesquisa: **é possível a automação do processo de criação de uma base de citações para os periódicos do SciELO?**

1.4 Objetivos

O objetivo geral deste trabalho consiste do desenvolvimento de uma metodologia para a criação de uma base de citações atualizada automática e continuamente a partir dos periódicos cadastrados no SciELO.

Os objetivos específicos se desdobram em: 1) descrever a metodologia proposta; 2) desenvolver um protótipo para testar a metodologia proposta; 3) validar o protótipo para testar a sua aplicação prática; 4) identificar fatores limitadores e propor alternativas para viabilizar a implantação e disponibilização da base de citações criada.

1.5 Premissas e limites da pesquisa

Como o ponto fulcral desta pesquisa é a proposta de uma metodologia, de um processo, não se pretende avaliar ou corrigir o conteúdo dos arquivos XML analisados, mas apenas garantir que esse conteúdo seja representado fidedignamente na base de citações criada.

Não foram usados critérios para filtrar as informações e, quando usados, foram explicitados ao longo do texto. Novos critérios podem ser introduzidos em análises futuras mais específicas.

Cabe ressaltar que os periódicos cadastrados no SciELO representam importância relativa para cada área de conhecimento, e não integra o escopo deste trabalho a análise específica de uma ou outra área. Eventuais análises superficiais foram utilizadas somente para ilustrar as possibilidades de recuperação de informações da base criada.

Conforme explicitado anteriormente, os passos dos processos descritos na FIG. 3 e na FIG. 5 garantem a atualização automática e contínua das informações necessárias à composição da base de citações. A premissa adotada neste caso é de que qualquer alteração – na lista de periódicos de uma Coleção ou nas estatísticas de um determinado periódico – será identificada no processo diário (desde que os *links* estejam acessíveis e que o padrão identificado não seja alterado), que por sua vez será tratado no processo de obtenção e interpretação de arquivos XML. Dado o volume de informações, cabe ressaltar que esse processo não é imediato.

A metodologia proposta depende da disponibilidade dos dados fonte e dos arquivos XML do SciELO, e no caso de não estarem disponíveis essas informações, outra solução automática ou semiautomática deverá prover esses dados.

1.6 Estrutura do trabalho

Este Capítulo tratou da contextualização e importância do tema proposto, apresentando as etapas necessárias à construção da base de citações do SciELO e de um resumo dos testes que foram realizados. O **Capítulo 2** discute a importância da produção e comunicação científica, a cientometria – especificamente com base no modelo inicial de criação da base de citações do ISI – e a importância da criação de índices de citações locais. Os exemplos apresentados sobre o uso de bases de citações para a realização de pesquisas mostram o potencial de exploração para a base de citação proposta. O **Capítulo 3** descreve a metodologia utilizada para o desenvolvimento deste trabalho, incluindo os passos metodológicos necessários à construção e validação do protótipo. O **Capítulo 4** detalha os resultados da primeira fase do protótipo, apresentando os dados cadastrais para as coleções tratadas. Os **Capítulos 5 e 6** discutem as particularidades do teste e da validação da segunda fase – a criação da base de citações – e os resultados obtidos para a revista *Perspectivas em Ciência da Informação* (Capítulo 5) e para a Coleção Saúde Pública (Capítulo 6). O **Capítulo 7** analisa os resultados encontrados e os principais problemas identificados. O **Capítulo 8** discorre sobre as conclusões e recomendações do trabalho, identificando as suas contribuições e sugerindo algumas possibilidades de realização de novas pesquisas.

2. PERIÓDICOS CIENTÍFICOS, COLEÇÕES DIGITAIS E CIENTOMETRIA

A apresentação dos fundamentos para a metodologia proposta neste trabalho inicia-se com a discussão sobre a importância da produção e comunicação científica, especificamente em periódicos, e os impactos da transição destes para o meio digital – as bibliotecas digitais. Em seguida discute-se a cientometria, parte da bibliometria que trata da literatura científica, especificamente com base no modelo inicial de criação da base de citações criada pelo ISI e na apresentação de alguns exemplos de exploração de suas informações. A importância da criação de índices de citações locais e alguns estudos cientométricos que se utilizaram do SciELO são discutidos. Finalmente, outra aplicação para uma base de citações é discutida: a avaliação de coleções digitais, especificamente em relação ao Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PORTAL CAPES).

2.1 Periódicos científicos e a internet

A importância da comunicação científica é enfatizada por Meadows já no prefácio de seu livro:

A comunicação situa-se no próprio coração da ciência. É para ela tão vital quanto a própria pesquisa, pois a esta não cabe reivindicar com legitimidade este nome enquanto não houver sido analisada e aceita pelos pares. Isso exige, necessariamente, que seja comunicada. Ademais, o apoio às atividades científicas é dispendioso, e os recursos financeiros que lhes são locados serão desperdiçados a menos que os resultados das pesquisas sejam mostrados aos públicos pertinentes. Qualquer que seja o ângulo pelo qual a examinemos, a comunicação eficiente e eficaz constitui parte essencial do processo de investigação científica. (MEADOWS, 1999, vii)

O autor descreve os primórdios da criação da revista científica como meio de comunicação, e atribui a Carlos II o patrocínio para a criação da *Royal Society* em Londres, em 1662, delineada de acordo com os últimos escritos de Francis Bacon sobre as atividades de uma instituição de pesquisa. Meadows explica que o modelo criado definia responsáveis pela coleta de informações em outros países para a *Royal Society*, e um deles – Denis de Sallo – começou a publicar notícias sobre o que acontecia na Europa no *Journal des Sçavans*, considerada a primeira revista em sentido moderno – datada de 1665.

Praticamente na mesma época surgia a *Philosophical Transactions*, primeira revista inglesa, porém com nítidas diferenças destacadas por Meadows e que fizeram com que fosse a inglesa a precursora do moderno periódico científico. “O *Journal des Sçavans*, por fim, verificou que era impossível manter o amplo leque de temas com que havia

começado e passou a se concentrar basicamente em temas não-científicos” (MEADOWS, 1999, p. 7). Conforme o autor, a partir da segunda metade do século XVII, o termo *journal* passou a ser aplicado para representar uma comunicação periódica formal de um conjunto de artigos.

Meadows (1999) destaca livros, periódicos científicos e congressos e conferências como os grandes canais de divulgação da pesquisa, mesmo considerados os diferentes graus de importância que cada um deles assume para áreas de conhecimento diferentes. Conforme o autor, entretanto, artigos de periódicos e livros são considerados as formas de publicação definitiva de resultados de pesquisas.

O autor apresenta três etapas do processo de comunicação científica: a primeira, em que os cientistas produzem informação; a segunda, na qual eles inserem a informação produzida em algum canal de divulgação, e a terceira, que corresponde à recuperação das informações nos diversos canais (MEADOWS, 1999, p.209). Para facilitar a terceira etapa, o autor comenta sobre o uso de índices e resumos como recursos auxiliares, e a importância da padronização da estrutura dos artigos para facilitar a recuperação, a exemplo do uso de palavras-chave representativas do conteúdo do trabalho.

Em relação à produção e comunicação do conhecimento científico, Cunha (2009, p. 25) afirma que:

A comunicação científica tem por finalidade divulgar o conhecimento produzido entre pesquisadores e profissionais de uma determinada área do conhecimento. Este conhecimento é produzido a partir de um objeto de estudo, ao qual se aplica um método de observação, percepção, formulação de questões, coleta de dados, análise e divulgação de resultados, denominado método científico. O método científico permite que o conhecimento gerado seja fruto de uma pesquisa que traga mais confiabilidade às questões observadas, razão pela qual se distingue do conhecimento popular, também chamado de senso comum.

A autora define as fontes primárias, secundárias e terciárias, e explica que “as fontes terciárias têm a finalidade de conduzir o usuário às fontes secundárias e primárias, como as bibliotecas, os centros de informações e os guias bibliográficos.” (CUNHA, 2009, p. 32).

O advento da *internet* facilitou a disseminação dos periódicos, oferecendo praticamente a mesma visibilidade a todas as revistas (MENECHINI e SILVA, 2012, informação verbal³). Paralelamente, a discussão sobre preços de periódicos e lucratividade das editoras – que envolveu entidades de prestígio como a Universidade de *Harvard* e a *Elsevier* – estimulou a disseminação de periódicos de acesso aberto, ideia lançada pelo

³ Entrevista realizada em 2012 com Rogério Meneghini, coordenador científico da SciELO, e Maurício da Rocha e Silva, editor da revista *Clinics*, publicação da Faculdade de Medicina da USP. Disponível em: <<http://www.youtube.com/watch?v=E-YI3c7JIng>>. Acesso em: 07/04/2013.

SciELO no Brasil em meados da década de 1990. Periódicos de acesso livre como a *Public Library of Science* (PLOS) são exemplos atuais dessa iniciativa. (MENEHINI e SILVA, *op. cit.*)

O crescimento da importância dos periódicos científicos como instrumento de avaliação dos programas de pós-graduação aumenta o interesse das instituições de ensino na manutenção e qualificação dos periódicos vinculados à organização, especialmente os de acesso livre e com arquivos abertos, e gera novas questões organizacionais cujos estudos apenas se iniciam. A própria CAPES recomenda que os periódicos adotem práticas e normas internacionais, indicando, inclusive, sistemas brasileiros de editoração eletrônica que obedeçam aos padrões internacionais, o que propicia o reconhecimento dos títulos e aumenta a visibilidade da produção científica brasileira. Nesse contexto, o Sistema Eletrônico de Editoração de Revistas (SEER) apresenta-se como uma das possíveis soluções para os programas de pós-graduação – ele foi desenvolvido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) em 2003 a partir da customização do *Open Journal System* (OJS) – *software* de gerenciamento e publicação de revistas eletrônicas desenvolvido pelo *Public Knowledge Project* (PKP), da *University of British Columbia*.

O SEER/OJS é um tipo de biblioteca digital de periódicos científicos, e o surgimento das bibliotecas digitais é destacado por alguns autores como uma transição das bibliotecas tradicionais. Mendonça (2006, p. 225) conceitua uma biblioteca tradicional e justifica o surgimento da biblioteca digital com o propósito de atender às necessidades de seus usuários:

Pode-se dizer que uma biblioteca tradicional é uma organização aberta, pois se acha inserida no meio ambiente que a cerca, influenciando-o e, ao mesmo tempo, sendo influenciada por ele, composta de funções e atividades relacionadas com a formação, desenvolvimento e organização de coleções (funções meio) e com a disseminação e recuperação da informação (funções fim), produzindo produtos e serviços que satisfaçam às necessidades informacionais de seus usuários. Em decorrência deste meio ambiente que se transforma cada dia com rapidez cada vez maior e com o propósito de atender às necessidades dos usuários, acompanhando a mutação ocorrida na sociedade, imposta pelos avanços científicos e tecnológicos, a biblioteca tradicional convive com o surgimento de outro tipo de biblioteca, a biblioteca virtual, conectada à rede e atendendo a uma gama de usuários superior à da biblioteca tradicional, haja vista não se prender limitações impostas pelo tempo e espaço.

Ao longo do texto, a autora apresenta a independência de horários e a não necessidade de deslocamento físico como fatores importantes para os usuários e abolidos nas bibliotecas digitais. Marcondes *et al.* (2006, p. 11) descrevem a transição das bibliotecas convencionais para as digitais da seguinte forma:

As bibliotecas sempre foram, historicamente, instituições que concentram informação em um lugar físico para servir a uma comunidade de usuários.

Como as bibliotecas eram físicas, o alcance de seus serviços ficava restrito às comunidades que a elas conseguiam ter acesso. Com o surgimento da Internet, esta situação evolui de forma drástica: não apenas o potencial de coletar e concentrar informações dispersas aumentou enormemente, como se tornou expressiva a capacidade de atender ao público no sentido mais amplo possível. As bibliotecas digitais tornam-se, desse modo, um instrumento poderoso de distribuição, cooperação e acesso ao conhecimento, atendendo e podendo servir de foco agregador a uma comunidade segmentada, distribuída geograficamente.

Ao adaptarem-se às novas necessidades na forma de biblioteca digital, as bibliotecas expandiram o número e a abrangência seus usuários, e conseqüentemente suas necessidades de informação.

Atualmente diversas bibliotecas digitais são gerenciadas a partir de informações de LOG. A esse respeito, especificamente sobre o acesso a periódicos eletrônicos, Dias (2002, p. 12) conclui que

mediante análise dos arquivos de log de acesso, não se pode ter um perfil completamente preciso do acesso a periódicos eletrônicos hospedados em sites da web, mas apenas um modelo aproximado do que acontece na realidade, pois esta é uma abordagem quantitativa que não fornece subsídios para endereçar questões de caráter qualitativo, tais como a opinião dos usuários com relação ao conteúdo do site, satisfação, usabilidade e os motivos que o levaram a acessar o site.

Os relatórios para gerenciamento das necessidades de usuários deveriam agregar outras informações, a exemplo do que sugere Cunha (2009, p. 21), ao apresentar um resumo a respeito dos estudos de comportamento de busca de usuários:

Estudos de comportamento de busca, em sua grande maioria, coletam informações sobre variáveis demográficas dos respondentes (tais como gênero, idade, área do conhecimento), fontes de informação preferidas, fontes de informação acessadas, frequência (sic) de acesso às fontes, necessidades de informação, tempo estimado para a busca, necessidade de auxílio na busca, independência nas buscas, material mais utilizado, dificuldades encontradas na busca, dentre outras.

Diversos autores apontam a falta de literatura específica a respeito das bibliotecas digitais. Entre eles, Marcondes *et al.* (2006, p. 11): “propõe-se este livro a atender à necessidade existente na literatura de ciência da informação no Brasil sobre bibliotecas digitais. Por ser a primeira iniciativa neste sentido, inscreve-se como uma contribuição ao conhecimento do assunto”. A partir das considerações apresentadas pelos autores consultados (MARCONDES *ET AL.*, 2006; CENDÓN e RIBEIRO, 2008; CUNHA, 2009), podemos extrair as seguintes considerações: o crescimento do uso e importância das bibliotecas digitais; pesquisas são geralmente mais restritas a áreas e segmentos específicos de usuários; existe pouquíssima literatura específica sobre a gestão de bibliotecas digitais.

Dentre os artigos consultados, destacamos o texto “Gestão em bibliotecas digitais” (VICENTINI, 2006). Ao estudarmos o referido texto, entretanto, notamos muito mais

uma preocupação com a forma das bibliotecas digitais do que propriamente com o seu conteúdo. O autor apresenta alguns trechos em que cita a preocupação com o atendimento às necessidades dos usuários de uma biblioteca digital, mas não se aprofunda na discussão sobre como isso pode ser realizado.

Especificamente em relação às bibliotecas digitais de periódicos científicos, diversas possibilidades de análise podem ser realizadas nos estudos cientométricos, tema apresentado a seguir a partir do uso da WoS.

2.2 Bibliometria e Cientometria: criação e uso das bases de citações

Os periódicos de uma área a princípio podem ser agrupados de acordo com a lei⁴ de Bradford, que de forma simplificada define três grupos – cada um com cerca de um terço do total de artigos pertinentes ao tópico específico nesses periódicos. O primeiro grupo, relativamente pequeno, possui grande concentração de artigos pertinentes; o segundo contém um número maior de periódicos moderadamente produtivos em termos de artigos pertinentes; e o último grupo, com o maior número de periódicos, apresenta artigos pertinentes ocasionais (MEADOWS, 1999, p.222).

Além da lei de Bradford, Alvarado (1984) cita também a Lei de Lotka, que descreve a produtividade dos autores; e a Lei de Zipf, que descreve a frequência no uso de palavras num determinado texto. No Brasil, de acordo com o autor, a bibliometria foi introduzida por Tefko Saracevic, Bert Boyce e Wilfred Lancaster, sendo o primeiro deles “o maior impulsor desta abordagem, tendo orientado 8 mestrados entre 1972 e 1975, incluindo-se entre estes, 3 dos 4 pioneiros da bibliometria no Brasil”.

Santos e Kobashi (2009) apresentam uma análise da constituição e institucionalização da bibliometria, da cientometria e da infometria. Conforme os autores, mesmo sem consenso sobre o uso desses termos, a bibliometria estaria mais associada à quantificação da produção científica para fins de gestão de acervos, a cientometria à interpretação desses dados (da produção em si, circulação e consumo da produção científica) com base em teorias das ciências humanas e sociais para auxiliar o planejamento e a avaliação de políticas científicas, e a infometria – utilizando-se das anteriores – à apreensão dos aspectos cognitivos da atividade científica. (SANTOS e KOBASHI, *op. cit.*, p. 155).

⁴ Na bibliometria a expectativa de comportamento da literatura é determinada “lei”

Em relação aos estudos de quantificação da atividade científica, os autores afirmam ser obrigatório o estudo de Lotka, Bradford, Zipf e Price. A partir deste último, que se valeu dos três primeiros para formular suas leis cientométricas, “os estudos quantitativos adquiriram novos contornos, centrando-se fundamentalmente na análise da dinâmica da atividade científica, incluindo tanto os produtos quanto os produtores de ciência” (SANTOS e KOBASHI, *op. cit.*, p. 158). Nesse sentido, de acordo com os autores, para Price a cientometria é o estudo quantitativo da atividade científica – um modelo pioneiro que integrou o estudo da atividade científica à bibliometria.

Essa ideia de Price seria a antecipação da ideia de construção de mapas da ciência que evidenciariam a importância de revistas científicas e sua cobertura (SMALL e GARFIELD, 1986), e esses mapas seriam o embasamento do projeto desenvolvido na década de 1970 pelo ISI (SANTOS e KOBASHI, 2009).

De acordo com Vanti (2011, p. 7), “os primeiros artigos em que o conceito de cientometria se fez presente foram publicados pelos pesquisadores do VINITI – *All-Union Institut for Science and Technical Information*, da Academia de Ciências da ex-URSS”. Conforme a autora, Nalimov (atraído pelas ideias da cibernética), Styazhkin e Vledutsiv (todos pertencentes aos quadros VINITI) publicam em 1959 o primeiro artigo considerado cientométrico: “As informações científicas e técnicas como uma das tarefas da cibernética”⁵. A autora comenta que somente dez anos depois deste trabalho, Nalimov e Mulchenko publicam a primeira monografia intitulada com o termo cientometria (*Naukometriya* em russo).

Apesar de a cientometria concentrar suas análises nos periódicos científicos, Santos e Kobashi (2009) alertam para o fato de que “não se pode reduzir a atividade científica à produção, à circulação e ao consumo de artigos de periódicos e, muito menos, confundir o crescimento quantitativo de artigos com o desenvolvimento cognitivo da ciência” (SANTOS e KOBASHI, *op. cit.*, p. 159).

Segundo Meneghini e Packer (2007), as publicações em cientometria e bibliometria de autores brasileiros cresceram exponencialmente. De acordo com Santos e Kobashi (2009), os estudos quantitativos da ciência tem crescido no Brasil, seguindo a tendência mundial. Além do crescimento do volume de publicações de autores brasileiros sobre cientometria e bibliometria, os autores comentam que também cresceram os estudos em nível de doutorado e mestrado, e as comunicações em congressos de diferentes campos de pesquisa.

⁵ Cabe lembrar que o termo “informatika” para os russos refere-se à informação científica.

Conforme Vanti (2011), Garfield compreendeu a relação entre o uso dos artigos de revisão na literatura periódica e as citações bibliográficas, sendo que as referências citadas em um artigo poderiam funcionar como elementos de recuperação de informação da mesma forma que as palavras-chave ou os descritores atribuídos por um indexador profissional. Para a autora, com base nessa ideia,

Garfield lança as bases para a criação do ISI em 1958, na Filadélfia. A partir de 1960, desenvolve novas ferramentas para facilitar o acesso às informações científicas de diferentes campos do conhecimento: Science Citation Index (SCI), Current Contents, Social Sciences Citation Index (SSCI) e Arts and Humanities Citation Index (A&HCI). (VANTI, 2011, p.9)

Esse conjunto de ferramentas, que recebeu o nome de *Web of Science*, é o tema do próximo tópico.

2.2.1 Web of Science

De acordo com dados do seu próprio site⁶, a *Web of Science*[®] cobre mais de 12.000 periódicos de grande impacto no mundo inteiro, além de mais de 150.000 conferências – em diversas áreas de conhecimento (mais de 250 disciplinas) a partir do ano de 1.900.

Mesmo que os índices disponíveis nessa base sejam criticados (MOTTA, 1983; ADAM, 2002), o ISI segue como “rei” absoluto, pois a grande maioria dos esforços para a construção de bases similares limita-se a uma ou poucas áreas de conhecimento, e não empreendem esforços para prover índices de citação nos moldes daqueles oferecidos pelo ISI.

Diversos autores ressaltam a importância da criação de uma base de citações para o desenvolvimento científico local, inclusive para a produção de políticas científicas nacionais (WEBSTER, 1998; XIN-NING, 2001; GOGOLIN *ET AL.*, 2003; NEGISHI, SUN e SHIGI, 2004; SULEIMENOV, 2009; GUIMARÃES, SILVA, SANTANA, BRAGA, BOCHNER e GOLDBAUM, 2011; CENDÓN, GUIMARÃES, SILVA, OLIVEIRA, MATTOS, SANTANA e FERNANDES, 2012; ZIBAVERA e PARMON, 2012). Para o desenvolvimento do protótipo proposto neste trabalho, interessa a esta pesquisa a forma como se iniciou a construção da base de citações do ISI, especificamente em relação às ideias de seu fundador Eugene Garfield.

Garfield (1972, p.527) explica que desde 1927 vários autores – entre eles Gross e Gross, e Bradford – mapeavam partes da rede existente de periódicos científicos, mas não

⁶ Disponível em: <<http://thomsonreuters.com/web-of-science/>>. Acesso em: 12 ago. 2013.

existia um mapa geral desses periódicos. Ele afirma que, apesar do interesse e esforço desses pesquisadores, a dificuldade prática para compilar o grande volume de dados de forma manual era o grande desafio a ser vencido.

Para romper esse desafio, o autor aponta como solução o uso dos dados disponíveis em meio magnético e usados para a produção do SCI, que havia passado de 600 periódicos em 1964 para 2400 em 1972 – e, no final de 1971, essa base continha mais de 27 milhões de referências. Outro trabalho do autor aponta para 15 milhões de artigos publicados desde 1945 e mais de 200 milhões de referências citadas (GARFIELD, 1992, p.2). Em 1995 ele afirma que existiam 3300 periódicos cadastrados (GARFIELD, 1995, p.88). O site do ISI, responsável pelo SCI, mostra que existem mais de 12.000⁷ periódicos atualmente.

De acordo com Garfield (1972, p.527-30), o ISI realizou em 1971 uma avaliação sistemática a partir dos dados armazenados, usando para isso uma amostra com os 3 últimos meses de 1969. Essa avaliação, gerada a partir de 2200 periódicos disponíveis no período, produziu três listagens a partir de aproximadamente 1 milhão de citações: a primeira acumulava o número de vezes que uma revista havia sido citada, e distribuía essas citações por ano em que foram citadas, conforme a FIG. 6 a seguir:

FIGURA 6 – Frequências de citações

ITEM NO	CITED JOURNAL	TOTAL	NUMBER OF TIMES CITED										
			1969	1968	1967	1966	1965	1964	1963	1962	1961	1960	REST
00243	ACTA PATH JAP	36	1	3	3	4	6	7	3	.	.	3	6
00244	ACTA PATH MICROBIOL	736	29	69	87	59	56	59	44	48	20	31	234
00909	AM J ANAT	637	7	27	37	56	32	41	15	26	15	21	360
00910	AM J BOT	1171	13	74	87	68	73	66	57	57	47	49	580
00911	AM J CANCER	103	103
00912	AM J CARDIOL	1238	73	201	199	247	134	70	78	66	53	73	44
03591	CAN J SOIL SCI	33	.	2	1	4	1	2	6	6	3	2	6
03592	CAN J SURG	61	2	6	4	3	13	11	3	3	5	1	10
03593	CAN J TECH	3	3
03594	CAN J ZOOL	356	46	38	40	28	24	20	19	29	17	22	73
08990	ISRAEL J AGR RES	29	1	1	7	.	1	8	7	2	1	.	1
08991	ISRAEL J BOT	16	.	.	3	5	4	3	1
08992	ISRAEL J CHEM	91	14	25	18	10	11	6	7
09651	J INVEST DERM	695	24	78	81	69	65	46	30	31	34	22	215
09652	J IOWA MED SOC	13	.	.	.	5	.	.	1	2	.	.	5
09653	J IRISH MED ASS	16	1	3	4	3	3	2
13390	P CALIF ACAD SCI	18	.	.	1	4	3	.	1	.	.	.	9
13391	P CAMBRIDGE PHIL SOC	389	8	22	23	11	12	9	13	11	17	3	260
19755	Z ANGEW CHEM	47	.	.	1	.	.	1	.	.	1	1	43
19756	Z ANGEW ENT	35	.	.	1	1	4	2	4	1	1	5	16
19757	Z ANGEW GEOL	49	2	7	5	8	5	5	4	4	2	1	6
19758	Z ANGEW MATH	10	1	.	1	.	1	.	1	1	1	.	4

Fig. 1. Journal citation frequencies. The data show the total number of times each journal was cited during the last quarter of 1969 and the distribution by publication date of the particular issues cited. The journals shown were taken from a list of more than 20,000 items (journals, books, reports, theses, and so on) cited during the last quarter of 1969 in journals covered by the SCI.

Fonte: Garfield, 1972, p.528

⁷ Disponível em: <http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science/>. Acesso em: 07 abr. 2013.

A segunda lista apresentava um detalhamento histórico de cada periódico citado, apresentando para cada um os periódicos citantes:

FIGURA 7 – Estatísticas dos periódicos citados

CITED JOURNAL CITING JRNL	TOTAL	NUMBER OF TIMES CITED										
		1969	1968	1967	1966	1965	1964	1963	1962	1961	1960	REST
J LINN SOC BOT	17*	0	3	0	3	1	0	0	0	1	1	8
NEW PHYTOL	5	0	1	0	1	0	0	0	0	1	0	2
ALL OTHER (10)	12	0	2	0	2	1	0	0	0	0	1	6
J LIPID RES	902*	33	108	109	98	133	121	75	83	60	58	24
BIOC BIOP A	83	5	8	16	6	13	15	4	6	5	1	4
BIOCHEM J	39	0	4	7	3	6	7	3	3	4	1	1
J BIOL CHEM	39	0	6	7	3	8	3	6	2	2	2	0
J CHROMAT	30	2	3	4	4	2	6	0	4	2	1	2
J CLIN INV	29	1	3	3	3	7	2	2	3	3	1	1
J LIPID RES	28	5	5	4	2	4	1	2	0	3	1	1
P SOC EXP M	25	0	3	1	3	4	3	2	5	1	1	2
MILIT MED	5	0	0	2	0	0	0	1	1	0	1	0
NY ST J MED	5	0	1	1	2	0	0	0	1	0	0	0
ALL OTHER (123)	219	6	19	16	29	43	27	23	15	15	23	3
J LOND MATH SOC	173*	16	19	19	11	8	10	2	8	4	8	68
J LOND MATH	71	8	12	7	5	4	4	0	4	2	4	21
T AM MATH S	11	2	1	2	1	0	0	0	1	0	0	4
P CAMB PHIL	8	1	0	3	0	0	0	0	0	0	0	4

Fig. 2. Statistics on cited journals. The data on cited journals show the total number of times each journal was cited in the last quarter of 1969 and the distribution by publication date of cited issues. For each cited journal, the figure identifies all other journals ("citing journals") that referred to it five times or more during the quarter year (and the distribution of these citations by publication date of cited issues). Journals that referred to the cited journal less than five times are grouped as "all other." The data were taken from a complete list of journals cited during the last quarter of 1969 in journals covered by the SCI.

Fonte: Garfield, 1972, p.529

A terceira e última lista produzida era similar à segunda, entretanto organizava os dados por periódico citante, detalhando os periódicos citados:

FIGURA 8 – Estatísticas dos periódicos citantes

SOURCE JOURNAL	REFERENCE JOURNAL	TOTAL	1969	1968	1967	1966	1965	1964	1963	1962	1961	1960	REST
J LIBR AUT	>	168*	15	50	35	19	8	14	5	6	3	1	12
PROGRAM		15	3	6	5	1	0	0	0	0	0	0	0
J LIBRARY AUTOMATION		7	3	4	0	0	0	0	0	0	0	0	0
ALL OTHER (120)		146	9	40	30	18	8	14	5	6	3	1	12
J LIPID RES	>	313*	15	36	35	26	25	29	20	10	15	15	87
J BIOL CHEM		43	2	6	3	6	1	3	1	0	1	1	19
J LIPID RES		28	5	5	4	2	4	1	2	0	3	1	1
BIOCHIM BIOPHYS ACTA		19	1	3	1	3	2	3	1	0	2	2	1
BIOCHEM J		13	1	1	3	0	0	0	1	2	1	1	3
J AMER CHEM SOC		12	0	0	0	0	0	0	0	0	1	0	11
BIOCHEMISTRY		9	0	3	0	2	1	2	1	0	0	0	0
J CHEM SOC LONDON		5	0	0	1	0	0	0	0	0	1	0	3
J CHROMATOGR		5	0	1	1	1	0	1	0	1	0	0	0
METHODS ENZYMOL		5	0	0	0	0	0	0	0	1	0	0	4
ALL OTHER (89)		124	5	12	19	10	8	16	10	2	5	6	31
J LOND MATH	>	743*	21	44	58	52	42	34	31	38	29	33	361
J LOND MATH		52	6	12	5	5	4	3	0	2	1	3	11
P LONDON MATH SOC		37	4	5	7	1	2	2	0	1	1	0	14

Fig. 3. Statistics on citing journals. The data on each citing journal ("source journal") show the total number of references each journal contained in the last quarter of 1969 and the distribution of that total by publication date of journal issues referred to. For each citing journal, the figure identifies all journals ("reference journal") cited five times or more during the quarter year (and gives the distribution by publication data of cited issue). Journals cited less than five times are grouped as "all other." The data are taken from a list of journals processed for the SCI during the last quarter of 1969.

Fonte: Garfield, 1972, p.530

Após discutir algumas particularidades das três listagens, Garfield (1972, p.540) afirma que os resultados possíveis a partir desse tipo de análise possuem um grande potencial de gerenciamento das coleções de periódicos, apontando como a principal aplicação da análise de citações a avaliação de pesquisas e políticas científicas. Ele conclui que o uso da base do SCI para mapear a rede de periódicos científicos pode contribuir para uma ciência mais eficiente.

Garfield (1979) discute também problemas relacionados ao uso da análise de citações como ferramenta de avaliação para medir o desempenho de pesquisas, pesquisadores e instituições, e afirma que há dois pontos básicos a serem tratados: o primeiro refere-se ao aumento da complexidade e do volume de informações científicas, o que exige um cuidado especial para criar critérios de avaliação que identifiquem as grandes

contribuições científicas; e o segundo relaciona-se à objetividade da análise de citações, e seu relativo baixo custo em relação a outras técnicas de pesquisa. Em outro trabalho (GARFIELD, 1992) o autor afirma que os dados quantitativos obtidos com a análise de citações não substituem nem reduzem a necessidade de uma análise qualitativa por parte de especialistas, oferecendo, na verdade, uma perspectiva única para complementar sua análise. Ele conclui que os dados de publicações e citações oferecem o potencial para a criação de indicadores objetivos para a avaliação de desempenho e que a maioria dos limites apontados na literatura sobre essa abordagem pode ser tratada estatisticamente.

Em relação à produção científica em países em desenvolvimento, Garfield (1995, p.88) aponta que a cobertura de periódicos no SCI é seletiva, apresentando periódicos de influência internacional – a representatividade de artigos de pesquisadores latino-americanos cresceu de 1% para 1,5% aproximadamente, entre 1970 e 1995 (10.000 dos 650.000 artigos indexados anualmente). O autor comenta que havia sugerido a criação de um SCI para a América Latina, porém desconhecia tal índice até então.

Após a análise de dados constantes no SCI sobre pesquisadores latino-americanos, Garfield (1995, p.94) apresenta uma série de recomendações, entre elas a de que a análise quantitativa da literatura oferece importantes contribuições para a definição de políticas de investimento e avaliação do desempenho de indivíduos, instituições e nações.

Alguns exemplos dessas contribuições são apresentados a seguir.

2.2.2 Aplicações práticas

É importante ressaltar que diversas bases de dados – além do ISI – oferecem recursos para análise cientométrica, entre elas a SCOPUS e o *Google Scholar* (GS). Os exemplos citados a seguir pretendem apenas ilustrar a diversidade de pesquisas que podem ser realizadas a partir de uma base de citações, e a importância da criação de bases que representem melhor a produção científica local para auxílio à definição de políticas e melhoria da qualidade dessa produção.

Winter, Zadpoor e Dodou (2013) discutem as diferenças entre a *Web of Science* e o GS. Enquanto o primeiro, não acessível ao público em geral, faz uma indexação seletiva a partir do argumento de que poucos periódicos contêm os artigos mais importantes e formam a base científica para todas as disciplinas, o GS é um serviço livre que coleta dados automaticamente. Essa coleta considera sites de periódicos, universidades e sites pessoais – e por esse motivo alguns autores discutem sua fragilidade, entre eles Jacsó (2005; 2008), Cathcart e Roberts (2005); Donlan e Cooke (2005); Wleklinski (2005) e Vine (2006).

Harzing (2013) apresenta a evolução do GS nas áreas de Química, Física, Economia e Medicina para o período 2012 e 2013, notando na comparação com o ISI que o GS apresenta uma boa cobertura para as últimas quatro décadas. Radicchi e Castellano (2013) usaram mais de 30.000 perfis de pesquisadores do GS – eles afirmam que análises com mais de 1.000 indivíduos eram raras, e que a disponibilidade de grandes bancos de dados como o do GS favorecem estudos de maior porte. Entretanto, Prathap (2013) apresenta possíveis falhas nos resultados de Radicchi e Castellano (2013) e sugere que amostras menores respondem melhor que o uso de “força bruta computacional” com uma amostra muito grande pode gerar falsas descobertas ou correlações espúrias.

Algumas aplicações de estudos realizados a partir do ISI incluem a colaboração científica, a partir da análise da coautoria (SCHUBERT e BRAUN, 1990). Kumar e Jan (2013) analisaram, a partir do ISI, a cooperação na produção científica entre a Turquia e a Malásia sobre o tema “combustíveis energéticos”. Meyer *et al.* (2013) usaram o ISI para analisar o desenvolvimento do tema “empreendedorismo”.

Em relação a estudos de gênero – outra importante possibilidade de exploração de uma base de citações – Nourmohammadi e Hodaei (2013) analisaram a produção científica feminina iraniana a partir do ISI. Hartley e Cabanac (2013) analisaram a preferência entre homens e mulheres no uso de gráficos e tabelas para a produção científica, e usaram como fonte os 752 periódicos publicados pela *Wiley-Blackwell* listados no JCR 2011, pois esses arquivos estavam disponíveis em HTML, facilitando assim a extração automática de informações. Sotudeh e Khoshian (2013) realizaram uma comparação entre gêneros na produção do JCR entre 2005 e 2007 na área de Nanotecnologia, e apontaram dificuldades para identificação do sexo de alguns autores. O Excel e o SPSS foram usados pelos autores para a análise dos artigos. Em suas conclusões, apesar da quantidade maior de publicações por homens, o impacto e produtividade não apresentaram diferença significativa entre os sexos.

Pesquisas locais usando o ISI incluem Sooryamoorthy (2009), que usou o SCI para analisar artigos em que ao menos um autor era sul-africano, enquanto Boshoff (2009) analisou a participação de autores de fora da África Central na produção de artigos em que ao menos um dos autores era filiado a instituições da África Central. Mênignbêto (2013) usou o ISI para analisar periódicos em que ao menos um autor era afiliado a instituições de um dos 15 países da África Ocidental. Exceto no caso da Nigéria, ele concluiu que há necessidade de maior integração entre esses 15 países para a produção científica.

Colazo-Reyes *et al.* (2008) analisaram os periódicos da América Latina e Caribe indexados no ISI de 1995 a 2003, e dividiram-nos em 3 grupos: locais, em que ao menos um dos autores é afiliado a instituição do mesmo país do periódico; regionais, quando ao menos

um dos autores é de outro país que aquele do periódico; e externos, quando todos os autores são de outro país. As conclusões mostraram que os periódicos de maior Fator de Impacto são os externos, que tendem a publicar em inglês, e que os periódicos locais apresentam tendência de publicação na língua local (português ou espanhol). González-Alcaide, Valderrama-Zurián e Aleixandre-Benavent (2012) argumentam que o uso do Fator de Impacto para a avaliação de periódicos em países que não adotam a publicação em inglês favoreceu o desenvolvimento de periódicos em inglês. Brody (2013) afirma que, apesar das críticas sobre o uso do Fator de Impacto como medida de avaliação, ele ainda é insubstituível.

A seguir discute-se a importância das bases de citações locais, usando o SciELO como exemplo.

2.3 Bases de citações locais e o uso do SciELO

A importância da criação de bases locais para uma melhor representação da produção científica é defendida por diversos autores. Negishi, Sun e Shigi (2004) descrevem a criação de um índice de citações para periódicos japoneses – o *Citation Database for Japanese Papers* (CJP) – e destacam a importância da criação de índices de língua não inglesa, a exemplo das bases chinesas *Chinese Science Citation Database*, *China Scientific and Technical Papers and Citations Database* e *Chinese Social Science Citation Index*. Conforme os autores, o CJP constitui-se em uma importante ferramenta para a comunidade acadêmica do Japão.

Zibareva e Parmon (2012) apresentam o *Russian Science Citation Index* e sugerem o uso dos dados bibliométricos gerados neste sistema para uma avaliação mais apurada dos pesquisadores e instituições locais. Suleimenov *et al.* (2009) descrevem, a partir da mesma justificativa da necessidade de uma base local, a criação do *Kazakh Science Citation Index*, que atualmente possui dados de aproximadamente 30.000 pesquisadores.

Ye (2013) também justifica a importância da criação de um banco de dados de citações para livros, e algumas possibilidades de aproveitamento dessas informações para a melhoria da qualidade de publicação dos mesmos e do desenvolvimento de coleções. O autor justifica a criação do banco de dados para preencher a lacuna local de informações desse tipo.

Em relação à América Latina e Caribe, Krauskopf *et al.* (1995) analisaram o ISI entre 1981 e 1993, para Argentina, Brasil, Chile, Colômbia, Costa Rica, Cuba, Jamaica, México, Peru e Venezuela. Collazo-Reyes (2013) demonstra que o português passou a ser

a segunda língua científica – atrás apenas do inglês – na produção científica registrada no ISI para países da América Latina e Caribe.

Buscando a análise bibliométrica da produção científica sobre Síndrome da Imunodeficiência Adquirida (SIDA) na América Latina e Caribe, Macias-Chapula, Rodea-Castro e Narvaez-Berthelenot (1998) usaram como fonte 4 bases: AIDSLINE (o banco de dados sobre SIDA da Biblioteca Nacional de Medicina dos Estados Unidos), o SCI, Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS) e PERIODICA (Literatura em Ciências na América Latina da Universidade Nacional do México). O uso de outras bases além do SCI é justificado pelos autores, pois sua cobertura não alcança boa parte dos periódicos de países em desenvolvimento. Os resultados confirmaram quantitativamente essa premissa, pois foram encontrados apenas 82 artigos no SCI (3,93%), contra 272 no PERIODICA (13,05%), 765 no LILACS (36,69%) e 966 no AIDSLINE (46,33%).

Meneghini, Mugnaini e Packer (2006) apresentam uma análise de 13 periódicos nacionais cadastrados no SciELO e listados no JCR. Os autores identificaram 2 grandes grupos de periódicos: os de orientação nacional, que seguem um padrão de citações (concedidas e recebidas) mais voltado para periódicos nacionais; e os de orientação internacional, cujas citações são em maioria relacionadas a periódicos internacionais. Em suas conclusões, os autores defendem a necessidade do uso de uma base local para sustentar políticas de financiamento e acompanhamento. O SciELO também foi usado por Macias-Chapula (2010) para analisar alguns periódicos de Saúde Pública: *Revista Médica de Chile*, *Archivos Latinoamericanos de Nutrición* (Venezuela) e *Salud Pública de Mexico*.

Em razão da importância do SciELO para a produção científica em países da América Latina e Caribe, um breve histórico dessa base é apresentado a seguir.

2.3.1 SciELO

Adam (2002) apresenta um interessante trecho de seu artigo com o título “Candidatos ao trono”. Além de colocar o ISI como rei, ele discute algumas iniciativas, entre elas o consórcio *Cross-Ref*, o *ResearchIndex* e o *Astrophysics Data System* da NASA – mas conclui que nenhum deles tem a cobertura de múltiplas áreas de conhecimento, sendo esse um fator importante que limita o acesso dessas iniciativas ao trono.

Em relação a esse critério, certamente o SciELO enquadra-se como um forte candidato – um príncipe, em potencial – por sua cobertura de diversas áreas de conhecimento e países, ressaltando-se o fato de que boa parte de seus periódicos não está indexada no SCI. A esse respeito, Mugnani (2006) concluiu que o uso de indicadores

provenientes do SciELO pode contribuir para a adequação de critérios usados na avaliação da produção científica nacional e para estimular um processo de melhoria da qualidade de periódicos nacionais. De acordo com o autor, atualmente esses critérios são muito direcionados a indicadores provenientes de bases internacionais, principalmente do ISI.

De acordo com Vasconcellos (1999), o SciELO é “o produto da Metodologia para a Preparação, Armazenamento, Disseminação e Avaliação de Publicações Científicas em Formato Eletrônico, desenvolvido com o principal objetivo de ampliar a visibilidade da produção científica brasileira”. Ainda de acordo com a autora, “a metodologia do Projeto é diferenciada do conjunto de informações levadas à Internet, por obedecer a rígido controle de avaliação, seleção e operacionalidade”.

Guedes (2012) apresenta o processo de criação do projeto SciELO com uma riqueza de detalhes muito grande, visto que partiu de entrevistas com inúmeros atores envolvidos no processo desde o seu início, entre eles Abel Packer (Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde – BIREME), Charles Pessanha (Associação Brasileira de Editores Científicos – ABEC), Rogério Meneghini (Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP), Hooman Momen (*Comisión Nacional de Investigación Científica y Tecnológica* – CONICYT, Chile) e Regina Célia Castro (BIREME). Mesmo atribuindo a criação do projeto a Abel Packer e Rogério Meneghini, Guedes (2012) reforça a importância da contribuição da BIREME e da FAPESP no processo.

Conforme o autor, o artigo “*Lost Science in the Third World*” - escrito por W. Wayt Gibbs e publicado em 1995 na *Scientific American* - apresentava uma ciência invisível produzida no terceiro mundo, o que “causou impacto em diversos pesquisadores no Brasil”, entre eles, o próprio Meneghini.

Guedes (2012, p.58-9) relata que, de acordo com Abel Packer, eram 3 os objetivos do SciELO:

O 1º era desenvolver uma metodologia buscando resolver a capacidade para publicação online, pois nessa época tanto Brasil quanto América Latina tinham poucas iniciativas nesse campo. A ideia era utilizar o estado da arte internacional para construir uma solução que movesse os periódicos brasileiros para a web. O 2º corroborando o ponto supracitado por Gibbs e Meneghini, era estabelecer um novo tipo de controle sobre os artigos, com contagem de citações construindo um índice na internet complementar ao desenvolvido pelo Institute for Scientific Information (ISI), construindo, assim, uma avaliação mais completa da produção científica brasileira. O 3º objetivo era montar o uso, ou seja, era medir através dos downloads o uso do SciELO. Isto, segundo Packer, era um passo lógico, visto que, os objetivos anteriores eram focados na visibilidade e acessibilidade do projeto.

O Projeto surgiu em janeiro de 1997, como resultado de convênio entre a FAPESP e a BIREME. (VASCONCELLOS, 1999; GOLDENBERG *ET AL.*, 2007; GUEDES, 2012). O trabalho foi realizado de maneira experimental no primeiro ano, e a partir do

segundo semestre de 1998 teve início a fase de expansão, ampliando o número de títulos de revistas nacionais. (VASCONCELLOS, 1999; GUEDES, 2012). A partir de 2002, passou a contar também com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GOLDENBERG ET AL., 2007; GUEDES, 2012).

Entre as referências à importância e impacto do SciELO, Gálvez (2006) analisou publicações biomédicas do Chile nas bases SCI, LILACS, *Medical Literature Analysis and Retrieval System Online* (MEDLINE) e SciELO, e concluiu que este último oferece a melhor opção para a divulgação local da produção científica. Prat (2000) afirma que o SciELO-Chile possibilitou o estabelecimento de um convênio com a MEDLINE, assegurando o acesso ao texto completo de suas revistas. A autora comenta que o INIST da França também mostrou interesse em incorporar o SciELO no seu site e em suas bases de dados especializadas, e mesmo o ISI mostrou interesse no projeto e facilitou o acesso a informações importantes para a análise do impacto das coleções.

O SciELO armazena dados estatísticos importantes para a análise de seus periódicos, e alguns exemplos de análises bibliométricas são listados a seguir.

Goldenberg *et al.* (2007) realizaram a coleta e interpretação manual dos indicadores bibliométricos e de uso das revistas científicas, disponíveis na seção "estatísticas" dos sites da coleção SciELO Brasil, tomando como exemplo a revista *Acta Cirúrgica Brasileira*. Solano e Valdivia (2003) realizaram um estudo sobre a SIDA, para um conjunto de revistas no SciELO, no período 1997 a 2003, obtendo os dados fonte de forma manual. Población e Goldenberg (2001) trabalharam dados bibliométricos do periódico *Acta Cirúrgica Brasileira*. Uma característica comum a todos esses trabalhos é a obtenção dos dados de forma manual.

Packer *et al.* (1998; 2001) descrevem a metodologia do projeto SciELO e as etapas necessárias à criação da base de dados, constituindo-se em um "conjunto de normas, guias, manuais, programas de computador e procedimentos operacionais que tem como fim a transformação dos textos de periódicos científicos para o formato eletrônico" (GUEDES, 2012, p.71).

De acordo com Packer (2011),

embora parte desse desempenho seja devido aos fatores clássicos que influenciam a prática das citações, como idioma de publicação, área temática e colaboração internacional, o aperfeiçoamento dos periódicos brasileiros requer políticas públicas que contribuam para superar suas limitações.

O autor conclui que o panorama dos periódicos brasileiros – composto por sua presença marcante, seu desempenho internacional limitado e a estrutura de caráter nacional que sustenta a sua editoração e publicação – evidencia progressos em política, gestão,

infraestrutura e capacidade de comunicação científica. O desafio principal que se apresenta, para ele, é minimizar

a assimetria que existe atualmente entre os títulos nacionais e internacionais no que se refere à penetração e impacto internacional. A superação desse desafio constitui um empreendimento nacional para o Brasil, que envolve o melhoramento do conjunto dos periódicos de qualidade de modo a posicionar os títulos com melhor desempenho como referências em suas respectivas áreas. As condições políticas e de infraestrutura estão dadas e os problemas a superar são conhecidos: a dispersão das instâncias de editoração e publicação, a persistência de amadorismo e corporativismo nas políticas, gestão e operação dos processos editoriais, a reduzida cooperação internacional na editoração dos periódicos e autoria dos artigos, a publicação predominante em português, a resistência de pesquisadores e gestores da ciência em valorizar os periódicos nacionais.

Entre as aplicações possíveis para o uso de uma base de citações local, a seguir discute-se o uso da bibliometria para avaliação do PORTAL CAPES, importante biblioteca digital do cenário brasileiro.

2.4 Outras aplicações

Dada a importância estratégica do PORTAL CAPES para a ciência nacional, a seguir é apresentada uma entre as possíveis aplicações de uma base de citações – a avaliação de coleções digitais de forma automática. Especificamente, discute-se uma possibilidade de uso da bibliometria para a gestão do acervo do PORTAL CAPES.

A avaliação de sistemas de informação em geral – e de coleções digitais – não é um tema simples. Uma das formas de se avaliar o seu sucesso é pela mensuração quantitativa do uso de seus periódicos para a produção científica. Essa mensuração pode auxiliar na avaliação da adequação de seu conteúdo à produção de conhecimento científico.

Conforme Lancaster (1996, p. 20):

Ao avaliar um acervo, o que se procura de fato é determinar o que a biblioteca deveria possuir e não possui, e o que possui, mas não deveria possuir, tendo em vista fatores de qualidade e adequação da literatura publicada, sua obsolescência, as mudanças de interesses dos usuários, e a necessidade de otimizar o uso de recursos financeiros limitados.

A presença na coleção, dos itens buscados, influencia o sucesso do serviço de uma biblioteca na satisfação dos interesses de seus usuários. Portanto, um índice importante na avaliação da adequação de coleções é a probabilidade de posse, que pode ser definida como “a probabilidade de se encontrar um item procurado pelo usuário no acervo da biblioteca” (LANCASTER, 1996, p. 17).

Esta mensuração da adequação de uma coleção a partir da probabilidade de posse pode ser realizada através de diversas técnicas da bibliometria, que

é o estudo dos aspectos quantitativos da produção, disseminação e uso da informação registrada. Usada pela primeira vez por Pritchard em 1969, a bibliometria desenvolve padrões e modelos matemáticos para medir esses processos, usando seus resultados para elaborar previsões e apoiar tomadas de decisão. (TAGUE-SUTCLIFFE, 1992, p. 1)

Glänzel (2003) apresenta uma breve história da bibliometria, definida inicialmente por Pritchard como uma aplicação da estatística e da matemática para livros e outras formas de comunicação, em 1969 – quase simultaneamente à definição de cientometria de Nalimov e Mulchenko. Conforme o autor, a cientometria seria uma aplicação mais restrita da bibliometria, direcionada apenas às comunicações científicas.

Entre os principais autores relacionados às origens dos estudos estatísticos da literatura científica, são listados Alfred J. Lotka com seu estudo pioneiro sobre a frequência de distribuição da produtividade científica, em 1926; Gross e Gross com seu estudo baseado em citações (apesar de ser considerado o primeiro estudo de análise de citações, não se enquadra no sentido atual do termo) para a seleção de periódicos a serem comprados na área de química para pequenas escolas, em 1927; e Bradford, em 1934. Esses estudos não foram aplicados até a década de 1960, conforme o autor, porque não se considerava a necessidade de investimento em sistemas com tal sofisticação estatística (GLÄNZEL, 2003, tradução do autor).

O interesse em relação à avaliação quantitativa da produção científica mudou radicalmente a partir de 1961, de acordo com Glänzel (2003), a partir do lançamento dos dois livros de Derek John de Solla Price: "*Science since Babylon*" e "*Little Science - Big Science*". Outros fatores importantes dessa época foram os avanços da tecnologia da informação e o crescimento da produção científica. Esse contexto favoreceu o desenvolvimento da bibliometria, e uma das principais fontes de pesquisa passou a ser o SCI.

É importante ressaltar que apesar de a análise bibliométrica ser criticada por alguns autores (CAMARGO JR, 2010), para o tratamento de grandes volumes de dados – que é o caso do SciELO – ela supera, em termos de uma relação entre custo e benefícios, outras formas mais rigorosas como a análise por pares (MUGNAINI, 2006).

Delanghe, Sloan e Muldur (2011) reforçam a importância dos estudos bibliométricos e afirmam que “a interação entre as necessidades da política, por um lado, e os avanços acadêmicos na área de bibliometria, por outro lado, foi fundamental na formação do desenvolvimento e utilização de indicadores bibliométricos” (tradução do autor).

Um exemplo prático do uso da avaliação de coleções é discutido a seguir, referente à análise do PORTAL CAPES.

2.4.1 Avaliação de coleções: o acervo do PORTAL CAPES

Exemplo do uso da bibliometria para a avaliação de coleções é o estudo de Fernandes e Cendón (2010). Os autores utilizaram a análise de citações – uma das técnicas de estudo bibliométrico – para avaliar a adequação do conteúdo do Portal de Periódicos da CAPES (PORTAL CAPES) em relação à produção de teses e dissertações defendidas no triênio 2005-2007 na Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG). O trabalho revelou, entre outros resultados, que 45% dos artigos citados estavam disponíveis na coleção do PORTAL CAPES e que o conjunto de dados parece indicar que há espaço para melhoria na coleção de CI no PORTAL CAPES. Os autores recomendam que a análise bibliométrica passe a fazer parte dos critérios adotados para a seleção de novos títulos do PORTAL CAPES.

O uso da bibliometria para a avaliação do PORTAL CAPES também é defendido por Correa *et al.* (2008), que afirmam que os seus efeitos não foram completamente mensurados ou qualificados, e que “são necessários estudos bibliométricos e cientométricos que avaliem as mudanças e o impacto que esse acesso trouxe à pós-graduação e à ciência brasileira”. (CORREA ET AL., 2008, p. 144)

Em razão da importância do PORTAL CAPES para a ciência nacional, a seguir é apresentada uma breve descrição dessa biblioteca digital de periódicos científicos e da forma como seu acervo vem sendo composto, para que ao final sejam discutidas possíveis contribuições deste trabalho aos estudos bibliométricos sobre o PORTAL CAPES – que podem contribuir para melhorar a gestão do seu acervo.

De acordo com Correa *et al.* (2008, p.129), “desde a segunda metade do século XVII, quando as primeiras revistas científicas passaram a existir, até o presente milênio, os padrões de comunicação entre os pesquisadores foram muito alterados”. As autoras descrevem as mudanças na forma de armazenamento e nos hábitos de recuperação e uso da informação, mas reforçam a importância do artigo de periódico como “o principal meio de comunicação na ciência, graças à legitimidade e credibilidade alcançadas mediante o sistema de avaliação por pares e a garantia da divulgação dos desenvolvimentos científicos recentes” (CORREA ET AL., 2008, p. 129). Nesse contexto, elas afirmam que as próprias revistas sofreram grandes transformações e que

os anos 1990 marcam o ápice daquilo que veio a ser conhecido mundialmente como a ‘crise dos periódicos’, ou seja, a incapacidade de as bibliotecas manterem as assinaturas das principais revistas científicas nas respectivas áreas, como resultado da escalada dos preços, impulsionada pelos editores comerciais que passaram a publicá-las e distribuí-las. (CORREA ET AL., 2008, p. 129)

Entre os modelos alternativos a este contexto, as autoras citam a criação do PORTAL CAPES em novembro de 2000 como um marco importante, “pois se constitui em uma biblioteca digital de informação científico-tecnológica atualizada e de qualidade, produzida em âmbito mundial, em textos completos” (CORREA *ET AL.*, 2008, p. 129). As autoras ainda afirmam que o PORTAL CAPES é o maior do mundo em capilaridade, atendendo todo o território nacional – e que os portais americanos maiores em volume à época (das Universidades de *Harvard* e *Massachusetts*) eram de acesso local.

O impacto de implantação do PORTAL CAPES foi tão grande que, conforme Correa *et al.* (2008), algumas editoras chegaram a mandar representantes para o Brasil com a finalidade de certificarem-se da existência da CAPES, uma vez que o país foi o primeiro da América Latina a negociar diretamente com as editoras. De acordo com as autoras, a CAPES foi homenageada em 2002, no lançamento de um portal argentino similar ao brasileiro e que contou com o auxílio da Agência para o seu desenvolvimento.

Alguns números atualizados a partir de informações da CAPES evidenciam a relevância e importância do PORTAL CAPES: uma coleção de mais de 30.000 títulos disponível para 320 instituições usuárias (COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR, 2012).

Fernandes (2012, p.33-41) destaca a importância do PORTAL CAPES para a produção científica nacional e traça um breve histórico sobre o mesmo, descrevendo de forma objetiva os fatores que antecederam a sua criação desde 1987 com a implantação do Programa de Aquisição Planificada de Periódicos para Bibliotecas Universitárias (PAP), do Programa de Apoio a Aquisição de Periódicos (PAAP) em 1995 e do sucesso alcançado pelo Programa Biblioteca Eletrônica (ProBE) – consórcio paulistano para a compra de periódicos eletrônicos – que foi o embrião do PORTAL CAPES lançado em 2000.

Ao retratarem o histórico, a evolução e a utilização do PORTAL CAPES ao longo dos seus primeiros 10 anos, Almeida, Guimarães e Alves (2010, p.220) destacam que

O Portal de Periódicos é um instrumento de política pública para subsidiar o acesso ao conhecimento científico, gerido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). Criado em 2000, no âmbito do Programa de Apoio à Aquisição de Periódicos Eletrônicos (PAAP), o Portal se constitui hoje em um dos maiores acervos mundiais nesse setor e é atualmente o principal mecanismo para o apoio bibliográfico às atividades de C, T & I no Brasil, o que garantiu uma base para os excepcionais avanços recentes da ciência brasileira.

Dentre os benefícios destacados pelos autores, a implantação do PORTAL CAPES permitiu a redução do custo médio do acesso por usuário, o direcionamento dos investimentos realizados pelo governo brasileiro na aquisição de periódicos, o acesso universal a uma coleção ampla e atualizada de periódicos internacionais e o preenchimento das enormes lacunas nas coleções das bibliotecas.

Durante esse primeiro decênio, os autores destacam que em 2003 ocorreu a reavaliação do conteúdo do PORTAL CAPES, quando alguns títulos foram excluídos por serem considerados pouco relevantes ou subutilizados. No mesmo ano, foi criada a Comissão de Negociação, responsável por buscar junto aos fornecedores e editores alternativas para a redução de custos dos contratos mantidos pela CAPES.

Entre as preocupações centrais para a gestão do PORTAL CAPES, conforme os autores, está a observância de uma faixa de custo aceitável para manutenção e expansão da sua coleção. Alguns fatores citados que possuem impacto direto nesse custo são a expansão dos programas de pós-graduação no Brasil – aumentando o número de instituições usuárias – e o reajuste e crescimento do valor das assinaturas praticado pelas editoras. Mesmo assim, o elevado investimento (cerca de US\$ 61 milhões em 2010) é muito inferior, de acordo com os autores, ao valor necessário para dotar todas as instituições usuárias com a mesma coleção.

A expansão do acesso ao PORTAL CAPES é apresentada pelos autores de acordo com o volume de acessos – de 3 milhões, em 2001, para 65 milhões em 2009 – e a sua importância para o desenvolvimento científico nacional é explicitada por eles quando concluem que

A análise dos dados no período 2001-2009 evidencia a evolução e a consolidação do Portal e demonstra claramente que sua utilização crescente, como instrumento de apoio às pesquisas realizadas, vem fortalecendo, com isso, a cultura de acesso a material bibliográfico online, que passa a se constituir ferramenta incorporada ao dia a dia da comunidade científica e acadêmica brasileira. (ALMEIDA, GUIMARÃES e ALVES, 2010, 242-3)

Entre as recomendações sugeridas pelos autores, destacam-se os seguintes aspectos em relação à coleção do PORTAL CAPES: (I) a ampliação “em consonância com a crescente demanda e a diversidade dos cursos de pós-graduação, de modo a estabelecer uma correspondência com a expectativa da comunidade. Dessa forma, há claramente a contínua necessidade de sua expansão” e (II) a necessidade de “um estudo sobre o acervo do Portal, seu conteúdo e sua organização, visando garantir seu uso racional, especialmente relacionado a custos e utilização”.

Em suma, é importante a expansão contínua da coleção atrelada à gestão eficiente dos investimentos e de seu uso. Entretanto, a partir das considerações apresentadas por alguns autores (MARCONDES *ET AL.*, 2006; CENDÓN e RIBEIRO, 2008; CUNHA, 2009), observa-se que apesar do crescimento do uso e importância das bibliotecas

digitais⁸, as pesquisas empíricas sobre sua utilização e impacto são restritas a áreas e segmentos específicos de usuários.

Especificamente em relação às pesquisas realizadas sobre o PORTAL CAPES destaca-se o trabalho de Cendón e Ribeiro (2008). As autoras analisaram 40 trabalhos de pesquisa sobre o PORTAL CAPES, apresentando um panorama sobre os focos, resultados e pesquisadores envolvidos com o tema. Algumas das conclusões apresentadas são discutidas a seguir:

- Existem poucas pesquisas sobre o tema PORTAL CAPES, e dos 40 (quarenta) trabalhos de pesquisa analisados pelas autoras apenas 11 (onze) estão focados no impacto do seu uso na pesquisa e na produtividade das instituições usuárias ou na adequação do seu conteúdo (CENDÓN e RIBEIRO, 2008, TAB. 11, p. 165);
- A maioria dos estudos focaliza uma ou duas instituições, geralmente aquela(s) de origem do(s) pesquisador(es);
- As áreas de conhecimento pesquisadas são restritas em relação à abrangência do material disponibilizado no PORTAL CAPES.

Entre os estudos mais recentes sobre o tema, Fernandes (2012) – ao buscar compreender o fenômeno da não utilização do PORTAL CAPES por docentes das Instituições Federais de Ensino Superior (IFES) brasileiras – comprovou estatisticamente, pelo teste do qui-quadrado, que algumas variáveis interferem no seu uso ou não-uso, a exemplo da faixa etária, da área do conhecimento, da universidade, do nível de formação, do tempo de docência e do tempo dedicado à pesquisa. Cabe ressaltar que o autor concluiu que

não se pode generalizar o comportamento em relação à utilização ou não do Portal pelas sub-áreas (sic) de uma mesma área do conhecimento, uma vez que a porcentagem de não utilização do Portal por docentes das diferentes sub-áreas (sic) de uma grande área do conhecimento divergiu bastante em relação a grande área. (FERNANDES, 2012)

Outro ponto destacado pelo autor é que “não ocorre variação de uso e não-uso (sic) do Portal por região geográfica brasileira, uma vez que a região que mais utiliza é a região Sul (85%) e a que menos utiliza é a região Nordeste (82,3%)” (FERNANDES, 2012). O estudo das áreas e subáreas de conhecimento também é um ponto importante analisado por uma série autores, entre eles Maia (2005), Cunha (2009), Martinez, Ferreira e Galindo (2011) e Fernandes (2012).

⁸ Assume-se que o PORTAL CAPES é uma biblioteca digital a partir da definição adotada por Fernandes e Cendón (2010, p. 323) e Correa *et al.* (2008), entre outros.

As pesquisas identificadas neste trabalho que tratam do estudo do uso do PORTAL CAPES podem ser divididas em dois grupos: um de avaliação do uso a partir da percepção dos usuários, que é coletada a partir da aplicação de questionários respondidos por pesquisadores usuários do PORTAL CAPES (MAIA, 2005; CUNHA, 2009; SANTANA e PEIXOTO, 2010; FERNANDES, 2012; MAIA e CENDÓN, 2012); e o segundo grupo, que parte da análise quantitativa das referências citadas e sua disponibilidade no PORTAL CAPES, sem o envolvimento direto dos usuários e independente de o material citado ter sido obtido no PORTAL CAPES (OLIVEIRA e ODDONE 2007; FERNANDES, 2009; FERNANDES e CENDÓN, 2010; SANTANA e PEIXOTO, 2010).

Para os dois grupos citados, é possível a identificação de quatro limitações na abrangência da análise dos estudos sobre o PORTAL CAPES: quantitativa e geográfica, em relação ao universo das instituições usuárias; e quantitativa e temática, em relação às áreas de conhecimento cobertas. Essas limitações devem-se, sobretudo, à necessidade de processamento manual das informações.

Almeida, Guimarães e Alves (2010, p. 228) destacam que para a composição inicial da coleção do PORTAL CAPES foram utilizadas as coleções em papel assinadas pelas instituições federais de ensino que recebiam recursos da CAPES, a coleção assinada pelo ProBE da FAPESP e as bases de dados referenciais indicadas pela Comissão Brasileira de Bibliotecas Universitárias (CBBU).

Em relação ao desenvolvimento da coleção no período 2000-2010, os autores relatam que os critérios inicialmente utilizados para expansão da coleção lidavam com a demanda direta dos coordenadores de área ou de pesquisadores e docentes de pós-graduação. Entretanto, ressaltam que esses critérios “estão hoje em desuso, uma vez que praticamente todo o acervo disponibilizado pelas editoras já está disponível no Portal” (ALMEIDA, GUIMARÃES e ALVES, 2010, p. 230). Em seguida, eles afirmam que “sabidamente, o crescimento do acervo é resultado da demanda” (ALMEIDA, GUIMARÃES e ALVES, 2010, p. 233), e que em 2010 as áreas de Ciências da Saúde, as Biológicas, as Ciências Exatas e da Terra e as Engenharias representavam 58% dos títulos, enquanto as Ciências Humanas, Sociais e Artes estavam contempladas com 38% dos títulos da coleção.

Para Fernandes e Cendón (2010, p.329), o modelo de avaliação da coleção do PORTAL CAPES poderia ter sua eficácia ampliada caso fosse adotado um procedimento de análise bibliométrica. Os autores avaliaram as contribuições de vários modelos usados para a manutenção e o desenvolvimento de coleções, entre eles Danilowicz e Szarski (1981) e Mueller (1991).

3. METODOLOGIA

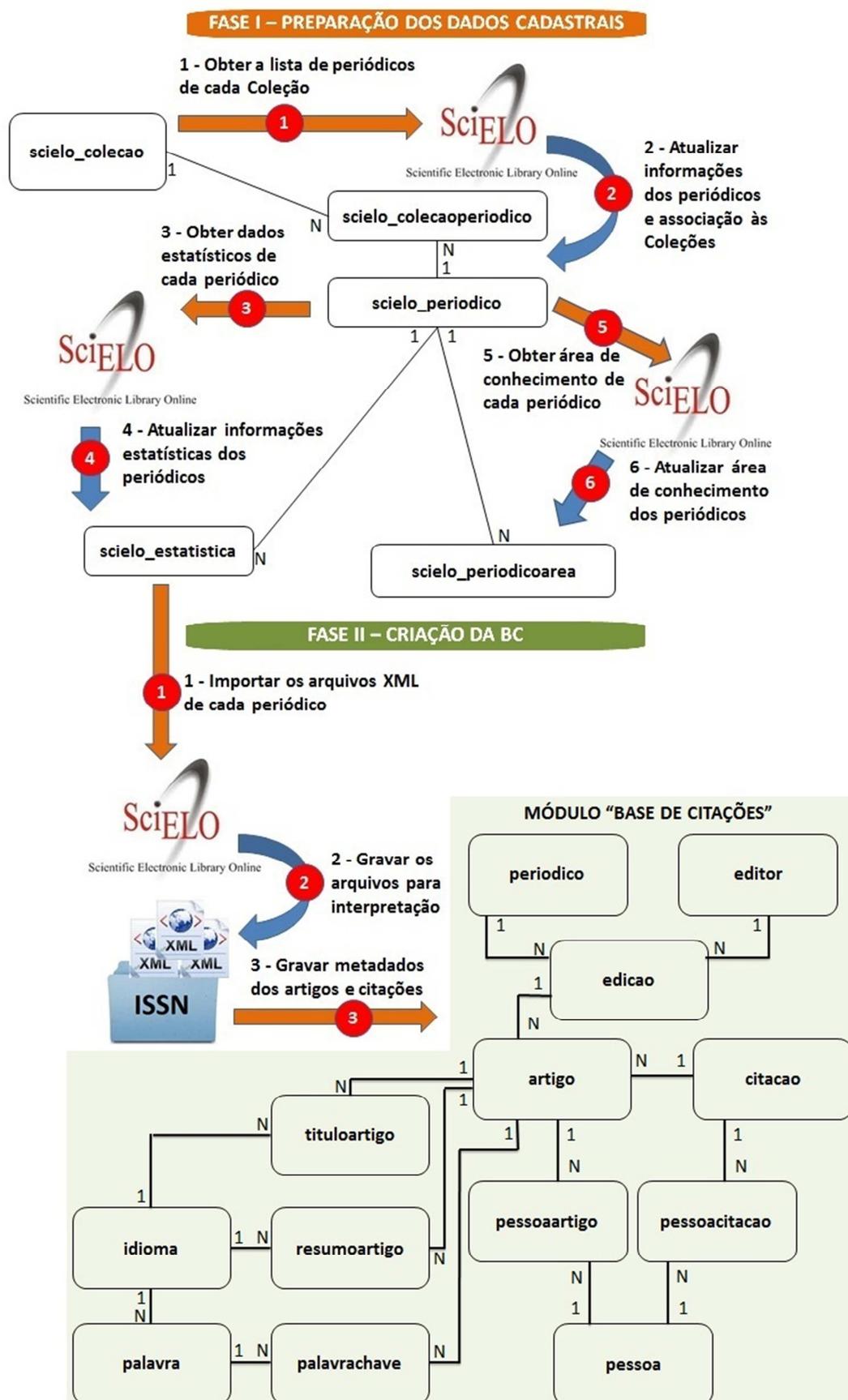
Este trabalho classifica-se como uma pesquisa aplicada (LAKATOS e MARCONI, 2007, p.20) ou exploratória, segundo Tripodi *et al.* (1975), que definem o propósito da pesquisa exploratória como a demonstração da viabilidade de um determinado programa ou técnica como uma solução em potencial para problemas práticos – objetivo central da criação da base de citações do SciELO.

Também pode ser classificada como híbrida, ao utilizar-se de métodos qualitativos e quantitativos (CRESWELL e CLARCK, 2011). Nesse sentido, trata-se de um trabalho qualitativo enquanto processo de desenvolvimento de um modelo relacionado à abstração, generalização e formação de conceitos para a metodologia proposta; e quantitativo, na medida em que o desenvolvimento do protótipo e da base de citações relacionam-se ao estudo objetivo da bibliometria. A combinação desses dois tipos pode proporcionar uma base contextual mais rica para interpretação e validação dos resultados, e de acordo com a definição dos autores esta pesquisa é interativa de prioridade qualitativa: interativa, pois os métodos qualitativo e quantitativo interagem e são integrados antes da análise final; a prioridade qualitativa refere-se ao propósito central de criação e abstração de um modelo para representação do processo de automação da construção da base de citações, e possui maior ênfase que a parte quantitativa, sendo esta última mais simples de acordo com a abordagem proposta.

A técnica usada na consecução dos passos metodológicos apresentados adiante tem como base – apesar de não seguir rigidamente suas notações – a modelagem relacional proposta por Codd (1969; 1970) e o Modelo Entidade-Relacionamento (CHEN, 1976; 2002). O uso desta técnica visa, em parte, a generalização do modelo construído para aplicação em outras coleções digitais.

A visão geral da metodologia completa é consolidada na FIG. 9 a seguir, que apresenta os dois passos e suas etapas, e o modelo relacional criado para os Módulos “Dados Cadastrais” e “Base de citações”:

FIGURA 9 – Visão geral da metodologia: passos e modelagem de dados simplificada



A seguir são descritos os passos metodológicos utilizados para o desenvolvimento deste trabalho, desde a escolha das ferramentas tecnológicas, passando pelo desenvolvimento dos experimentos e criação de sistemas, e chegando até a validação final do protótipo.

3.1 Definição das ferramentas para desenvolvimento

Para o desenvolvimento do protótipo e realização dos experimentos descritos foi usado o MySQL, um sistema de gerenciamento de banco de dados (SGBD) com base na *General Public License* (GPL), e que apresenta uma fácil integração com a linguagem de programação PHP, além de ser multiplataforma (funciona tanto no sistema operacional Windows como no sistema operacional Linux), e ter excelente desempenho e estabilidade. (GUIMARÃES, SILVA, SANTANA, BRAGA, BOCHNER e GOLDBAUM, 2011).

O ambiente de desenvolvimento dos experimentos apresenta a seguinte configuração: sistema operacional *Windows 7 Home Premium, service pack 1*, 64 bits; editor de PHP *Zend Studio 5.0.0* e *Zend Guard 4.0.0* para criptografia dos programas a serem disponibilizados na internet; *SQLyog 7.02* para manipulação do banco de dados MySQL. No ambiente *web* funcionam os programas PHP desenvolvidos, criptografados com a ferramenta *Zend Guard* e transmitidos com a ferramenta *FileZilla*; o banco de dados MySQL é administrado a partir do uso da ferramenta *PHPMyAdmin* em ambiente Linux.

Os navegadores *Chrome* e *Firefox* foram usados ao longo do desenvolvimento dos experimentos e desenvolvimento de sistemas, sempre atualizados com a versão mais recente. Os dois navegadores foram usados aleatoriamente, tanto no ambiente de desenvolvimento quanto no ambiente *web*.

3.2 FASE I: Obtenção dos dados do SciELO

As 4 etapas descritas adiante tratam da obtenção dos dados do SciELO: 1) identificação dos periódicos; 2) obtenção dos dados estatísticos anuais de cada periódico; 3) vinculação da(s) área(s) de conhecimento; e 4) criação do Módulo “Dados Cadastrais” do banco de dados. No teste do protótipo com a PCI, o *International Standard Serial Number* (ISSN) foi informado manualmente na primeira etapa, e as etapas 2, 3 e 4 foram realizadas automaticamente. Para a validação final com a Coleção Saúde Pública do SciELO, os dados dos 15 periódicos foram identificados automaticamente desde a primeira etapa.

3.2.1 Identificação dos periódicos de uma Coleção do SciELO

Cada coleção do SciELO apresenta uma página com a lista alfabética completa de seus periódicos correntes e não correntes. No caso da Coleção de Saúde Pública, o *link* http://www.scielosp.org/scielo.php?script=sci_alphabetic&lng=pt&nrm=iso apresenta esta relação, conforme a FIG. 10:

FIGURA 10 – Relação de periódicos da Coleção Saúde Pública



Lista Alfabética - 15 periódicos listados

Títulos correntes - 15 periódicos listados

- [Annali dell'Istituto Superiore di Sanità](#) - 15 números
- [Bulletin of the World Health Organization](#) - 163 números
- [Cadernos de Saúde Pública](#) - 222 números
- [Ciência & Saúde Coletiva](#) - 93 números
- [Gaceta Sanitaria](#) - 71 números
- [MEDICC Review](#) - 7 números
- [Revista Brasileira de Epidemiologia](#) - 57 números
- [Revista Cubana de Salud Pública](#) - 25 números
- [Revista de Salud Pública](#) - 48 números
- [Revista de Saúde Pública](#) - 263 números
- [Revista Española de Salud Pública](#) - 98 números
- [Revista Panamericana de Salud Pública](#) - 189 números
- [Revista Peruana de Medicina Experimental y Salud Pública](#) - 14 números
- [Salud Colectiva](#) - 27 números
- [Salud Pública de México](#) - 129 números

Fonte: SciELO, 2013⁹

Para a obtenção automática da relação de periódicos da Coleção foi desenvolvido um programa que interpreta o conteúdo do *link* acima, listado na FIG. 10, e extrai a situação (corrente ou não corrente), o título, o ISSN e a quantidade de fascículos de

⁹ Disponível em: <http://www.scielosp.org/scielo.php?script=sci_alphabetic&lng=pt&nrm=iso>. Acesso em: 04 ago. 2013.

cada periódico para armazenamento no banco de dados. Este programa analisa a referida página diariamente, mantendo atualizada a lista de periódicos.

Identificou-se nos *links* do SciELO, para o acesso a múltiplas coleções, uma pequena variação no *link* de acesso à lista de periódicos, com o padrão a seguir:

http://DOMÍNIO/scielo.php?script=sci_alphabetic&lng=pt&nrm=iso,

no qual o DOMÍNIO é apresentado no QUADRO 1 para cada Coleção. É importante observar que a única exceção a este padrão foi a Coleção de Periódicos da Bolívia, que apresenta o final **[/scielo.php/script_sci_alphabetic/lng_pt/nrm_iso](#)**:

QUADRO 1 – Informações sobre as Coleções do SciELO: domínios

COLEÇÃO	DOMÍNIO	LINK da lista de periódicos		
África do Sul	www.scielo.org.za	http://	www.scielo.org.za	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Argentina	www.scielo.org.ar	http://	www.scielo.org.ar	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Bolívia	www.scielo.org.bo	http://	www.scielo.org.bo	/sciELO.php?script=sci_alphabetic/ing_pt/nrm_iso
Brasil	www.scielo.br	http://	www.scielo.br	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Brasil <i>Proceedings</i>	www.proceedings.scielo.br	http://	www.proceedings.scielo.br	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Chile	www.scielo.cl	http://	www.scielo.cl	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Colômbia	www.scielo.org.co	http://	www.scielo.org.co	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Costa Rica	www.scielo.sa.cr	http://	www.scielo.sa.cr	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Cuba	sciELO.sld.cu	http://	sciELO.sld.cu	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Espanha	sciELO.isciii.es	http://	sciELO.isciii.es	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
México	www.scielo.org.mx	http://	www.scielo.org.mx	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Paraguai	sciELO.iics.una.py	http://	sciELO.iics.una.py	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Peru	www.scielo.org.pe	http://	www.scielo.org.pe	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Portugal	www.scielo.gpeari.mctes.pt	http://	www.scielo.gpeari.mctes.pt	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Saúde Pública	www.sciELOsp.org	http://	www.sciELOsp.org	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
<i>Social Sciences</i>	socialsciences.sciELO.org	http://	socialsciences.sciELO.org	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Uruguai	www.sciELO.edu.uy	http://	www.sciELO.edu.uy	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
Venezuela	www.sciELO.org.ve	http://	www.sciELO.org.ve	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso
<i>West Indian Medical Journal</i>	caribbean.sciELO.org	http://	caribbean.sciELO.org	/sciELO.php?script=sci_alphabetic&lng=pt&nrm=iso

Fonte: Desenvolvido pelo autor

3.2.2 Obtenção dos dados estatísticos anuais de cada periódico

Outras informações importantes apresentadas pelo SciELO estão contidas em uma lista de dados fonte para os periódicos indexados, a exemplo dos dados a seguir¹⁰:

FIGURA 11 – Lista de dados fonte: Revista Peruana de Medicina Experimental y Salud Pública

REVISTA PERUANA DE
MEDICINA EXPERIMENTAL
Y SALUD PÚBLICA

Revista Peruana de Medicina Experimental y
Salud Pública
ISSN 1726-4634

Data do último
processamento
10-03-2013

Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev Peru Med Exp Salud Publica	12	278	6509	296	23.17
2012	4	100	2025	83	25.00
2011	4	102	2334	104	25.50
2010	4	76	2150	109	19.00
total	12	278	6509	296	23.17

Fonte: SciELO, 2013

Para o acesso a essa página de dados fonte de cada periódico, foi identificado o padrão:

[http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=scl&issn=ISSN&CITED%5B%5D=TITULO&YNG%5B%5D=all](http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=scl&issn=<u>ISSN</u>&CITED%5B%5D=<u>TITULO</u>&YNG%5B%5D=all)

Dessa forma, foi desenvolvido um programa para acessar, interpretar e armazenar os dados estatísticos anuais de cada periódico a partir de seu ISSN e título. Para o caso da PCI, foi usado o *link* com seu ISSN e título:

[http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=scl&issn=1413-9936&CITED\[\]=PERSPECTIVAS+EM+CIENCIA+DA+INFORMACAO&YNG\[\]=all](http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=scl&issn=<u>1413-9936</u>&CITED[]=PERSPECTIVAS+EM+CIENCIA+DA+INFORMACAO&YNG[]=all)

¹⁰ Disponível em: <

http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=spa&issn=1726-4634&CITED%5B%5D=REVISTA+PERUANA+DE+MEDICINA+EXPERIMENTAL+Y+SALUD+PUBLICA&YNG%5B%5D=all>. Acesso em: 26 mar. 2013.

Os dados obtidos sobre cada periódico são atualizados diariamente e gravados no banco de dados, conforme detalhamento no tópico descritivo do Módulo “Dados Cadastrais”, adiante.

Para a importação automática das informações de Coleções do SciELO, levantou-se que a identificação das estatísticas de cada Coleção também segue um padrão determinado, que inclui, além do ISSN e do título do periódico, um PREFIXO associado a cada Coleção e apresentado no QUADRO 2 adiante:

[http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=PREFIXO
&issn=ISSN&CITED%5B%5D=TITULO&YNG%5B%5D=all](http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=PREFIXO&issn=ISSN&CITED%5B%5D=TITULO&YNG%5B%5D=all)

A exceção a este padrão é a Coleção de Periódicos da África do Sul, que apresenta pequena diferença no *link* de acesso às estatísticas:

http://statbiblio.za.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=PREFIXO&issn=ISSN&CITED%5B%5D=TITULO&YNG%5B%5D=all

QUADRO 2 – Informações sobre as Coleções do SciELO: prefixos

COLEÇÃO	PREFIXO
África do Sul	sza
Argentina	Arg
Brasil	Sci
Chile	chl
Colômbia	col
Cuba	cub
Espanha	esp
México	mex
Portugal	org
Saúde Pública	spa
Venezuela	ven

Fonte: Desenvolvido pelo autor

3.2.3 Vinculação da(s) área(s) de conhecimento de cada periódico

O SciELO também disponibiliza um *link* para cada Coleção com a lista de periódicos organizada por assunto – no caso, por área de conhecimento. Para a Coleção de Saúde Pública, o *link* http://www.scielosp.org/scielo.php?script=sci_subject&lng=pt&nrm=iso fornece essa lista. Outro programa foi criado para interpretar o conteúdo deste *link* e armazenar no banco de dados a(s) área(s) de conhecimento de cada periódico identificado na etapa anterior, realizando o processo diariamente para manter os dados atualizados.

De acordo com o levantamento realizado, concluiu-se que o acesso às áreas de conhecimento de cada periódico também segue o mesmo padrão identificado no QUADRO

1, variando apenas uma parte do *link*, alterando-se o parâmetro *script=sci_alphabetic* para *script=sci_subject*:

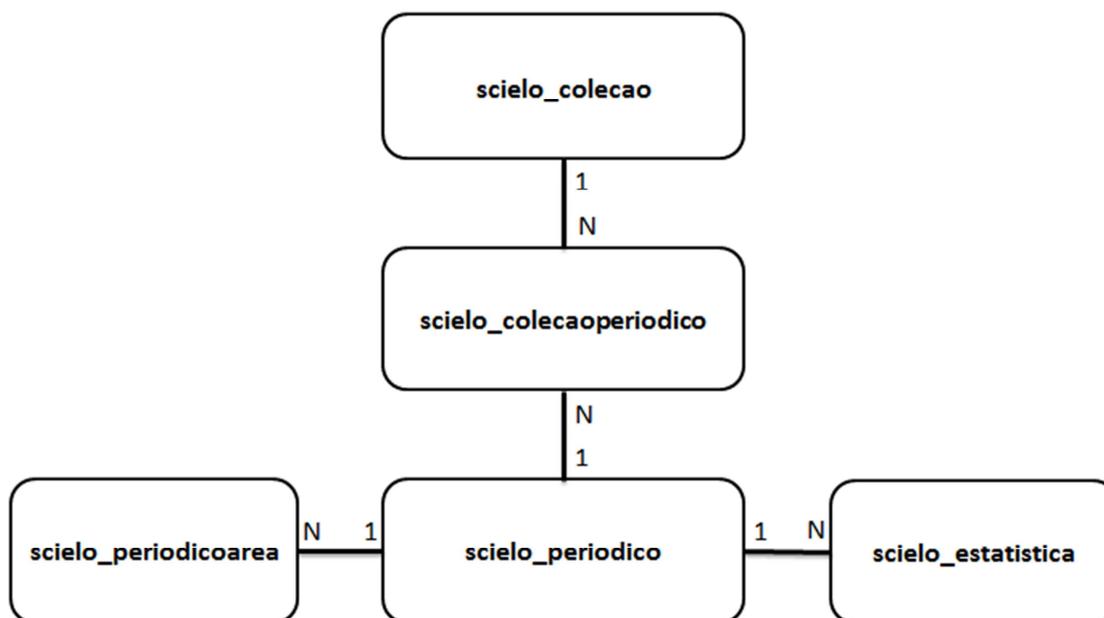
http://DOMÍNIO/scielo.php?script=sci_subject&lng=pt&nrm=iso,

Assim, a partir da informação sobre o DOMÍNIO de cada Coleção, é possível identificar-se a sua lista de periódicos e também as respectivas áreas de conhecimento. É importante ressaltar que o parâmetro *lng=pt* garante a padronização dos nomes das áreas em português, permitindo posteriormente o agrupamento de periódicos de diferentes Coleções por área de conhecimento.

3.2.4 Criação do Módulo “Dados Cadastrais” do banco de dados

Os dados cadastrais identificados nas etapas anteriores foram armazenados em um banco de dados, especificamente no Módulo “Dados Cadastrais”. A FIG. 12 apresenta o modelo relacional simplificado desse Módulo:

FIGURA 12 – Módulo “Dados Cadastrais”



Fonte: desenvolvida pelo autor

A tabela *scielo_periodico* armazena os dados de cada periódico identificado: ISSN, título, total de fascículos e situação (corrente ou não corrente). A tabela *scielo_periodicoarea* armazena a relação entre cada periódico e a(s) área(s) de conhecimento à(s) qual(is) ele está vinculado. Os dados estatísticos anuais de cada periódico (número de fascículos, total de citações concedidas e recebidas, e a média de artigos por fascículo) são armazenados na tabela *scielo_estatistica*. A tabela *scielo_colecao*

armazena o nome da coleção e os *links* de acesso à lista de artigos, às estatísticas e aos arquivos XML. A relação entre cada coleção e seus periódicos é armazenada na tabela *scielo_colecao*.

É importante ressaltar que a única tabela que tem seus dados informados manualmente é *scielo_colecao* – todas as outras são preenchidas a partir de dados coletados de forma automática e contínua.

A partir dessa análise, conclui-se que para a obtenção automática da lista de periódicos, das estatísticas e dos arquivos XML de uma Coleção do SciELO, é preciso identificar-se o DOMÍNIO e o PREFIXO de cada Coleção. Essas informações permitem a definição de um critério objetivo para a escolha das Coleções que comporão a amostra: coleções do SciELO que apresentem o *link* para acesso à lista de periódicos, às estatísticas e aos arquivos XML. Coleções sem um PREFIXO ou DOMÍNIO explicitados, no caso, não serão consideradas para a composição da amostra por ser impossível a identificação automática dos *links* para acesso à lista de periódicos, estatísticas ou arquivos XML. Essas informações são apresentadas a seguir:

QUADRO 3 – Informações sobre as Coleções do SciELO

COLEÇÃO	DOMÍNIO	PREFIXO	OBSERVAÇÃO
África do Sul	www.scielo.org.za	sza	
Argentina	www.scielo.org.ar	arg	
Bolívia	www.scielo.org.bo	-	SEM PREFIXO
Brasil	www.scielo.br	scl	
Brasil <i>Proceedings</i>	www.proceedings.scielo.br	-	SEM PREFIXO
Chile	www.scielo.cl	chl	
Colômbia	www.scielo.org.co	col	
Costa Rica	www.scielo.sa.cr	-	SEM PREFIXO
Cuba	scielo.sld.cu	cub	
Espanha	scielo.isciii.es	esp	
México	www.scielo.org.mx	mex	
Paraguai	scielo.iics.una.py	-	SEM PREFIXO
Peru	www.scielo.org.pe	-	SEM PREFIXO
Portugal	www.scielo.gpeari.mctes.pt	org	
Saúde Pública	www.scielosp.org	spa	
<i>Social Sciences</i>	socialsciences.scielo.org	-	SEM PREFIXO
Uruguai	www.scielo.edu.uy	-	SEM PREFIXO
Venezuela	www.scielo.org.ve	ven	
<i>West Indian Medical Journal</i>	caribbean.scielo.org	-	SEM PREFIXO

Fonte: Desenvolvido pelo autor

A partir dessas informações definiu-se inicialmente um conjunto com 11 Coleções de Periódicos: África do Sul, Argentina, Brasil, Chile, Colômbia, Cuba, Espanha, México, Portugal, Saúde Pública e Venezuela. Posteriormente, como a Coleção da Argentina não retornou nenhum arquivo XML, a mesma foi retirada da população. O resumo dos dados dessas 10 Coleções foi apresentado na FIG. 4.

3.3 FASE II: Criação da Base de Citações

A segunda Fase proposta trata da criação da base de citações, e para esse processo são necessários três passos: o primeiro corresponde à identificação e armazenamento dos arquivos XML disponíveis no SciELO contendo os artigos dos periódicos pertencentes às coleções; o segundo, à interpretação desses arquivos para extração dos metadados e informações sobre cada referência citada; e o último, que deve armazenar todas as informações extraídas de cada arquivo XML no banco de dados – no Módulo “Base de Citações”. Esses passos são descritos a seguir.

3.3.1 Levantamento dos arquivos XML de cada periódico

Para a obtenção dos arquivos XML do SciELO o primeiro passo foi a identificação do padrão de composição do *link* de acesso a cada arquivo contendo os artigos dos periódicos pertencentes à coleção a ser trabalhada. Esse padrão identificado diferencia-se pelo ISSN de cada periódico consultado, o ano de publicação, o número do fascículo e o sequencial do artigo:

<http://www.scielosp.org/scieloOrg/php/articleXML.php?pid=IDENTIFICADOR&lang=pt>

onde o IDENTIFICADOR corresponde a

S9999-9999999999999999
ISSN ANO NUM SEQ

Assim, S0102-311X2013000700019 significa:

ISSN = 0102-311X

ANO = 2013

NUMERO = 0007

Sequencial dentro do número = 00019 (19º arquivo)

E o acesso para o arquivo é possível a partir do *link*:

<http://www.scielosp.org/scieloOrg/php/articleXML.php?pid=S0102-311X2013000700019&lang=pt>

A partir das informações armazenadas no Módulo “Dados Cadastrais” para cada periódico, especificamente em relação às estatísticas anuais, torna-se possível a identificação automática dos anos para os quais existem arquivos XML, quantos fascículos existem em cada ano e quantos artigos (arquivos) estão disponíveis. A partir dessas informações foi criado um programa que gera os *links* conforme o padrão acima e captura

os arquivos XML – preservando como nome do arquivo o IDENTIFICADOR descrito acima – e armazena-os em um arquivo compactado e nomeado com o ISSN do periódico.

Para o acesso a múltiplas coleções, a identificação completa do *link* para o arquivo XML é dada pela expressão já descrita anteriormente, que também depende da informação sobre o DOMÍNIO:

<http://DOMÍNIO/scieloOrg/php/articleXML.php?pid=IDENTIFICADOR&lang=pt>

3.3.2 Interpretação dos arquivos XML do SciELO

A estrutura dos arquivos XML do SciELO apresenta 2 grandes grupos de informações: dados gerais sobre o artigo, e dados específicos sobre cada referência utilizada. O QUADRO 4 apresenta as principais *tags* identificadas e o tipo de informação armazenada em cada uma delas. As *tags* listadas estavam inseridas em dois grandes grupos: um contido nas tags <front> e </front> que apresenta dados gerais do artigo, como título, periódico, volume, edição, páginas, palavras-chave e resumos; e outro contido nas tags <back> e </back> com o detalhamento de cada referência citada:

QUADRO 4 – Estrutura do arquivo XML do SciELO

TAG	DESCRIÇÃO
<front>	Contém os metadados gerais do artigo
<journal-meta>	Apresenta o ISSN, título e título abreviado do periódico, e o nome do editor
<article-meta>	Contém os dados específicos do artigo: doi; título em cada idioma; nome e sobrenome dos autores; instituição dos autores; resumo em cada idioma; palavras-chave; dia, mês e ano de publicação; volume, número e páginas
<back>	Apresenta os dados de cada referência citada
<ref id="Bn">	Cada referência é agrupada dentro de uma <i>tag</i> identificada com um número n sequencial. Estão disponíveis informações sobre o tipo de citação; nome e sobrenome dos autores; título e idioma; fonte; dia, mês e ano de publicação; volume e número; páginas; editor e local.

Fonte: Desenvolvido pelo autor

A interpretação dessas *tags* permitiu a separação dos metadados de cada arquivo e de cada citação. A FIG. 13 apresenta o resultado parcial da extração de citações de um arquivo XML do SciELO:

FIGURA 13 – Resultado parcial da interpretação do XML SciELO: citações de um artigo

label	citation-type	article-title-pt	article-title-en	article-title-es	collab	source	year	month	day	publisher-name	publisher-loc	conf-name	conf-loc	conf-date	page-range	volume	issue	numero
1					COMUNIDAD ANDINA	Régimen común sobre derecho de autor y derechos conexos	1993											
2						Estudio sobre las limitaciones y excepciones al derecho de autor en beneficio de bibliotecas y archivos (Comité Permanente de Derecho de Autor y Derechos Conexos, OMPI)	2008											
3	book					Towards consensus on the electronic use of publications in libraries	2001			SUB	Göttingen							
4					EIFL	Draft law on copyright. Including model exceptions and limitations for libraries and consumers. Based on WIPO draft law on copyright and related rights (version 2005)	2009											
5	journal			Derecho de autor y bibliotecas digitales: a la búsqueda del equilibrio entre intereses contrapuestos		Transinformação	2008								123-131	20	2	2
6	journal			Protección tecnológica y privilegios de las bibliotecas: regulación en la legislación de derecho de autor de los países de la Unión Europea		Nuovi Annali della Scuola Speciale per Archivisti e Bibliotecari	2009								225-240	23		
7	book			Bibliotecas y derechos de autor: análisis comparativo de la nueva legislación de España y Portugal		Información, investigación y mercado laboral en información y documentación en España y Portugal	2008			Universidad de Salamanca	Salamanca				801-811			

Fonte: desenvolvida pelo autor

Para o primeiro teste, os arquivos XML obtidos no passo anterior para a PCI para o ano de 2012 foram interpretados por um programa específico que, ao final do processamento, apresentou um resumo para validação com os mesmos dados estatísticos da base SciELO: total de fascículos, de artigos, a média de artigos por fascículo e o total de citações.

FIGURA 14 – Resumo da importação automática de arquivos XML: PCI 2012

Arquivos processados: 34. Tempo total: 20.846111 segundos. Médio: 0.613122

ANO	VOLUME	EDICOES	ARTIGOS	MEDIA ARTIGOS	CITACOES
2012	17	3	34	11.3333	1038
TOTAL	-	3	34	11.3333	1038

Fonte: desenvolvida pelo autor

Em relação à interpretação dos arquivos XML, não foram necessários experimentos para a resolução de problemas de uso de diferentes normas para elaboração de referências, tais como a norma da Associação Brasileira de Normas Técnicas (ABNT) ou *Vancouver*. A importação de arquivos XML nesses formatos não gerou problemas de inconsistências, pois a estrutura de *tags* permaneceu inalterada. A partir dos testes realizados, infere-se que o SciELO faz o tratamento das diferenças de formatos antes da geração do arquivo XML, não havendo necessidade de tratamento desse tipo de problema.

Em relação à classificação dos materiais em cada citação, foi possível a identificação dos tipos *book*, *journal* e *confpro*, e para os demais casos essa informação não estava preenchida. Alguns exemplos são listados na FIG. 15:

FIGURA 15 – Tipos de citações encontradas: amostra

tipocitacao	nomefonte	ano
journal	Persp. Ci. Inf.	jan.
book	Epistemologia: curso de atualização	1980
confpro	Anais...	2003
	A face oculta do documento: tradição e inovação no limiar da Ciência da	2009

Fonte: desenvolvida pelo autor

Conforme explicitação anterior, não é objetivo central deste trabalho a geração de dados bibliográficos, e por isso algumas sugestões ou problemas identificados – como o tipo de citação em branco e o ano “jan.” na FIG. 15 – não foram tratados. As sugestões seguem ao final do trabalho como contribuição deste.

O primeiro exemplo, ilustrado na FIG. 16, refere-se ao item “A face oculta do documento: tradição e inovação no limiar da Ciência da Informação” que é uma tese, mas não está classificada com esse tipo – uma oportunidade de melhoria na geração do arquivo XML, que foi devidamente conferido:

FIGURA 16 – Possível melhoria para a geração do arquivo XML: identificação de teses

```

<ref id="B62">
<nlm-citation citation-type="">
<person-group person-group-type="author">
<name>
<surname><![CDATA[RABELLO]]></surname>
<given-names><![CDATA[R.]]></given-names>
</name>
</person-group>
<person-group person-group-type="editor">
<name>
</name>
</person-group>
<source><![CDATA[A face oculta do documento: tradição e inovação no limiar da
Ciência da Informação]]></source>
<year>2009</year>
<page-range>331p</page-range></nlm-citation>
</ref>

```

Fonte: desenvolvida pelo autor

O segundo exemplo, apresentado na FIG. 17, sugere que os periódicos com informação de mês textual parecem não ser corretamente identificados para a geração dos arquivos XML:

FIGURA 17 – Possível erro na geração do arquivo XML: datas textuais

```

<ref id="B4">
<nlm-citation citation-type="journal">
<person-group person-group-type="author">
<name>
<surname><![CDATA[ARBOIT]]></surname>
<given-names><![CDATA[A. E.]]></given-names>
</name>
<name>
<surname><![CDATA[BUFREM]]></surname>
<given-names><![CDATA[L. S.]]></given-names>
</name>
<name>
<surname><![CDATA[FREITAS]]></surname>
<given-names><![CDATA[J. L.]]></given-names>
</name>
</person-group>
<article-title xml:lang="pt"><![CDATA[Configuração epistemológica da Ciência da
Informação na literatura periódica brasileira por meio de análise de citações (1972-
2008)]]></article-title>
<source><![CDATA[Persp. Ci. Inf.]]></source>
<year>jan.</year>
<month>/a</month>
<day>br</day>
<numero>1</numero>
<issue>1</issue>
<page-range>18-43</page-range><publisher-loc><![CDATA[v. 15 ]]></publisher-loc>

```

Fonte: desenvolvida pelo autor

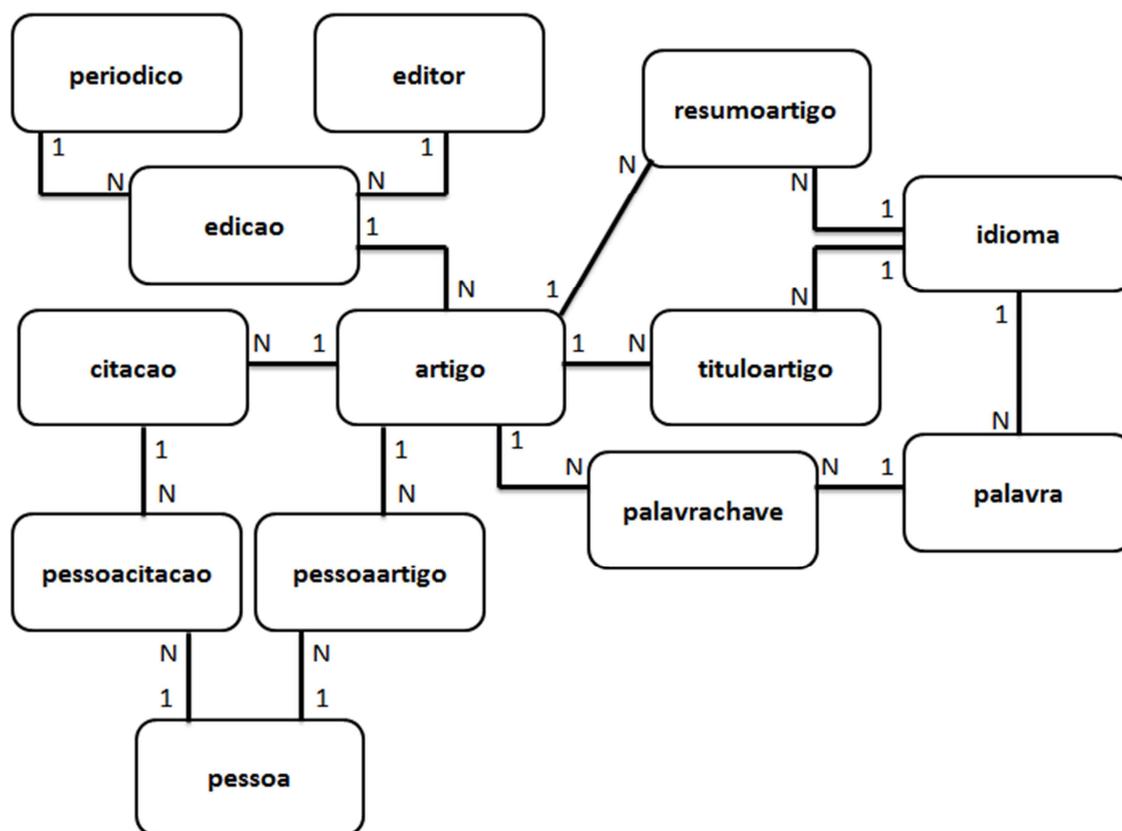
O programa que interpreta os arquivos XML armazena as informações sobre cada artigo (arquivo XML) e suas citações no banco de dados cuja estrutura é apresentada no tópico 3.3.3 a seguir.

3.3.3 Criação do Módulo “Base de Citações” do banco de dados

Após a identificação detalhada dos dados passíveis de extração do XML, foi modelado o banco de dados para armazenar a base de citações. O desenvolvimento desse modelo foi revisado a cada novo arquivo XML importado, pois eventualmente outros campos foram identificados e o banco de dados foi adequado para o seu correto armazenamento.

Os metadados identificados na etapa anterior foram armazenados especificamente no Módulo “Base de Citações”. A FIG. 18 apresenta o modelo relacional simplificado desse Módulo:

FIGURA 18 – Módulo “Base de Citações”



Fonte: desenvolvida pelo autor

A tabela *periodico* armazena o título do periódico e seu ISSN; a tabela *editor* armazena o nome do editor, e a tabela *edicao* relaciona *editor* e *periodico*. Cada artigo vincula-se a uma única *edicao*, e na tabela *artigo* são armazenados o identificador e o *Digital Object Identifier* (DOI) de cada arquivo importado. O título e o resumo de cada artigo são armazenados nas respectivas tabelas (*tituloartigo* e *resumoartigo*) com a identificação do idioma (títulos em português, espanhol, inglês, francês, africanês e latim). A tabela *idioma*

também identifica cada *palavra* encontrada na lista de palavras-chave dos artigos – os artigos são ligados às palavras a partir da tabela *palavrachave*.

Cada artigo armazenado possui várias citações, cujos metadados estão separados em campos específicos da tabela *citacao*. Existe uma tabela (*peessoa*) com todos os nomes e sobrenomes de pessoas identificadas, tanto nos metadados dos artigos como das citações – essas pessoas são relacionadas a cada artigo (*peessoaartigo*) ou citação (*peessoacitacao*).

3.4 A seleção da amostra

A escolha da PCI para o teste inicial deveu-se à facilidade de acesso ao banco de dados da mesma, liberado pelos seus editores e pessoal técnico. Além disso, a revista era a única representante do estrato Qualis A1 da área de Sociais Aplicadas I da CAPES e única revista da área de CI indexada no SciELO à época. O volume de artigos, ou arquivos XML, foi considerado satisfatório para os testes iniciais: 292. Os resultados desse teste são apresentados no Capítulo 5.

Para a validação do protótipo, em razão de projeto de pesquisa em parceria entre a ECI-UFMG e a Fundação Oswaldo Cruz (FIOCRUZ), foi utilizada a Coleção Saúde Pública do SciELO, com o total de 23.780 arquivos XML. Os resultados da validação são apresentados no Capítulo 6.

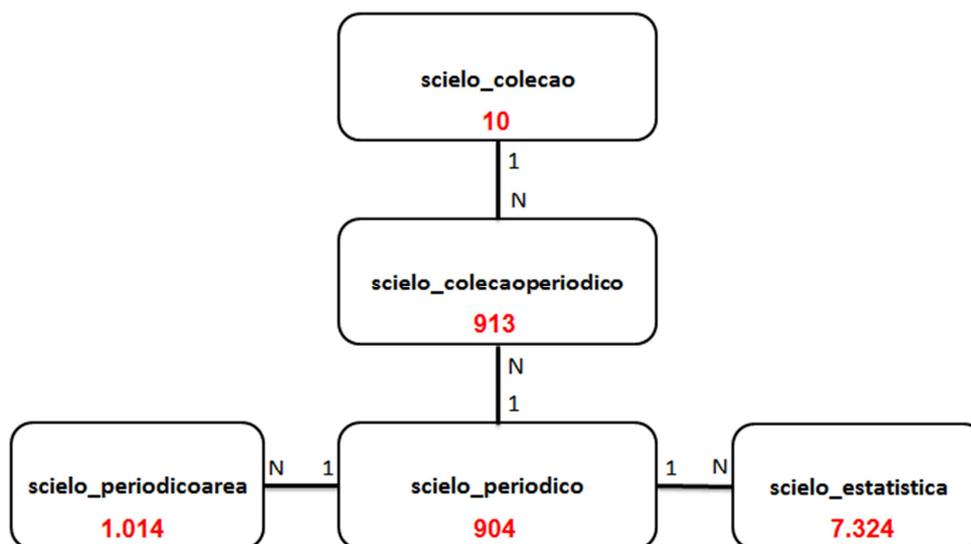
4 RESULTADOS DA FASE I: COLEÇÕES DO SCIELO EM NÚMEROS

O teste da primeira fase avaliou os resultados encontrados para a obtenção automática dos dados cadastrais do SciELO, e seus resultados são apresentados a seguir. Como não existe uma interface visível para o usuário, pois o processamento do sistema apenas grava as informações capturadas da internet no banco de dados, os totais de registros gravados em cada tabela do Módulo "Dados Cadastrais" são apresentados. Em seguida, descreve-se um importante limite identificado nesta fase em relação ao conteúdo dos dados fonte do SciELO: dados divergentes para um mesmo ISSN, em duas coleções distintas. Finalmente, é apresentada uma visão das coleções investigadas por área de conhecimento - uma das possibilidades para a realização de novas pesquisas a ser discutida no último capítulo.

4.1 Resumo dos dados estatísticos: o Módulo Dados Cadastrais em números

Os dados importados do SciELO em relação aos dados fonte são apresentados quantitativamente a seguir, em relação ao total de registros para cada tabela do banco de dados:

FIGURA 19 – Módulo "Dados Cadastrais" em números



Fonte: desenvolvida pelo autor¹¹

¹¹ Consulta realizada ao banco de dados em 18 jun. 2013.

Foram importadas as informações estatísticas de 10 coleções (*scielo_colecao*), totalizando 904 periódicos distintos (considerado o ISSN) armazenados em *scielo_periodico*. Considerando cada ano para o qual foram importadas as estatísticas foram encontrados 7.324 registros (*scielo_estatistica*). Dos 904 periódicos, 9 pertencem a duas coleções, totalizando 913 periódicos em *scielo_colecaoperiodico* (quando consideradas as 9 repetições). Em relação às áreas de conhecimento, considerando os periódicos que apareceram vinculados a mais de uma área, foram encontrados 1.014 registros em *scielo_periodicoarea*. O detalhamento das áreas de conhecimento é apresentado no tópico específico adiante.

4.2 Identificação dos periódicos e obtenção dos dados estatísticos

Foram identificados 904 periódicos (ISSN distintos) para as 10 coleções analisadas. Os totais de periódicos correntes e não correntes foram conferidos nas respectivas páginas do SciELO manualmente, para todas as coleções, e não foram identificadas diferenças na data da realização das consultas. A maior parte dos periódicos identificados (895) apresenta-se vinculado a uma coleção somente, e apenas 9 periódicos foram encontrados em 2 coleções, sendo uma delas sempre a Coleção Saúde Pública. A TAB. 1 apresenta esses periódicos, bem como os totais de fascículos, artigos, citações concedidas e recebidas para cada coleção – quando há diferença entre números das duas coleções para um mesmo ISSN, eles são destacados em vermelho:

TABELA 1 – Periódicos encontrados em mais de uma Coleção do SciELO

ISSN	COLEÇÃO	FASCÍCULOS	ARTIGOS	CITAÇÕES CONCEDIDAS	CITAÇÕES RECEBIDAS
0036-3634	México	98	1.014	35.064	4.761
	Saúde Pública	127	1.279	43.387	2.875
0124-0064	Colômbia	57	644	15.759	1.135
	Saúde Pública	47	575	13.949	302
0213-9111	Espanha	71	866	24.963	3.411
	Saúde Pública	68	826	22.552	2.019
0864-3466	Cuba	60	666	12.279	1.610
	Saúde Pública	25	334	7.147	347
1135-5727	Espanha	97	811	25.235	2.040
	Saúde Pública	96	805	23.697	971
1413-8123	Brasil	91	2.435	75.377	7.951
	Saúde Pública	90	2.405	74.338	2.618
0102-311X	Brasil	220	3.990	116.679	28.050
	Saúde Pública	220	3.991	116.703	13.581
0034-8910	Brasil	263	3.511	67.724	25.292
	Saúde Pública	263	3.511	67.720	11.756
1415-790X	Brasil	57	672	19.600	2.861
	Saúde Pública	57	672	19.600	1.392

Fonte: desenvolvida pelo autor

Como se observa na TAB. 1, os 9 periódicos apresentam algum tipo de divergência em relação ao total dos dados estatísticos – dados obtidos exclusivamente do SciELO. Todas as consultas aos dados fonte de cada ISSN em cada coleção estão disponíveis no Apêndice A.

Cada uma dessas situações é detalhada nas TABELAS 2 a 10 – uma para cada ISSN – destacados em vermelho todos os anos que apresentaram alguma diferença. Para cada ano, são apresentados os dados da Coleção Saúde Pública (CSP) e da outra coleção à qual o ISSN está vinculado. Os dados divergentes são apresentados em vermelho.

O primeiro ISSN analisado, 0036-3634, apresenta os dados entre 1997 e 2013 – e existem divergências em todos os anos. De 1997 a 2000, e no ano de 2013, somente existem informações disponíveis para a Coleção Saúde Pública:

TABELA 2 – Dados fontes do SciELO: diferenças para o ISSN 0036-3634

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	México	CSP	México	CSP	México	CSP	México
1997	6	-	67	-	2.152	-	84	-
1998	6	-	55	-	1.694	-	101	-
1999	8	-	68	-	2.106	-	123	-
2000	6	-	52	-	1.252	-	68	-
2001	6	6	61	61	1.933	1.933	67	94
2002	7	7	78	78	2.409	2.409	78	110
2003	11	11	120	120	3.802	3.802	186	229
2004	6	6	45	45	1.496	1.496	124	263
2005	6	6	46	46	1.472	1.472	145	266
2006	8	8	85	85	2.832	2.832	227	426
2007	10	10	97	97	3.117	3.117	244	416
2008	10	10	104	108	3.963	3.966	285	593
2009	10	10	119	119	5.088	5.088	333	624
2010	8	8	93	93	2.963	2.963	276	597
2011	10	10	101	101	3.769	3.769	246	576
2012	7	6	70	61	2.529	2.217	219	567
2013	2	-	18	-	810	-	69	-
	127	98	1.279	1.014	43.387	35.064	2.875	4.761

Fonte: desenvolvida pelo autor

Para o período 2001 a 2012, em todos os anos foram identificadas divergências em relação ao total de citações recebidas. Para o ano de 2008 as divergências também foram encontradas entre o total de artigos e citações concedidas. No ano de 2012 todas as informações – número de fascículos, artigos, citações concedidas e recebidas – apresentaram divergência entre as duas coleções.

O segundo ISSN, 0124-0064, apresentou diferenças entre as duas coleções (Coleção Saúde Pública e Coleção Colômbia) em todos os anos consultados: entre 2001 e 2012:

TABELA 3 – Dados fontes do SciELO: diferenças para o ISSN 0124-0064

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	Colômbia	CSP	Colômbia	CSP	Colômbia	CSP	Colômbia
2001	-	3	-	17	-	466	-	12
2002	-	4	-	25	-	427	-	20
2003	1	3	5	18	118	482	4	8
2004	4	4	22	22	660	660	11	16
2005	3	3	28	28	646	646	13	25
2006	5	5	52	52	1.003	1.246	21	53
2007	4	4	54	54	1.172	1.168	19	61
2008	6	6	88	88	2.190	2.190	47	114
2009	6	6	89	89	2.208	2.208	42	145
2010	8	8	100	100	2.625	2.605	67	202
2011	6	6	87	87	2.099	2.115	47	239
2012	4	5	50	64	1.228	1.546	31	240
	47	57	575	644	13.949	15.759	302	1.135

Fonte: desenvolvida pelo autor

Todos os anos apresentaram diferença em relação às citações recebidas. Entre 2001 e 2003 as informações para a Coleção Saúde Pública não estavam disponíveis. As informações de citações concedidas são divergentes para os anos de 2006, 2007, 2010 e 2011. Em 2012, todas as informações apresentaram diferenças.

Outro ISSN consultado, o 0213-9111, apresentou diferença em todos os anos (2002 a 2013) em relação às citações recebidas:

TABELA 4 – Dados fontes do SciELO: diferenças para o ISSN 0213-9111

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	Espanha	CSP	Espanha	CSP	Espanha	CSP	Espanha
2002	6	6	58	58	1.509	1.507	98	107
2003	7	7	72	73	1.838	1.843	110	128
2004	7	8	93	102	2.609	2.784	95	188
2005	6	6	58	58	1.527	1.527	133	172
2006	6	6	63	63	128	1.521	20	267
2007	6	6	67	67	1.694	1.639	137	182
2008	6	6	79	79	2.339	2.339	250	352
2009	6	6	93	93	3.136	3.136	304	447
2010	6	6	78	78	2.383	2.383	293	453
2011	6	6	80	80	2.638	2.678	298	516
2012	6	6	85	85	2.751	2.751	281	490
2013	-	2	-	30	-	855	-	109
	68	71	826	866	22.552	24.963	2.019	3.411

Fonte: desenvolvida pelo autor

Os anos de 2004 e 2013 apresentaram divergência em todas as informações. Além disso, foram identificadas diferenças do total de citações concedidas para os anos de

2002, 2003, 2006, 2007 e 2011. O ano de 2003 apresentou também divergências no total de artigos.

O quarto ISSN analisado, 0864-3466, apresentou diferença nas informações de todos os anos disponíveis no SciELO: 1995 a 2013:

TABELA 5 – Dados fontes do SciELO: diferenças para o ISSN 0864-3466

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	Cuba	CSP	Cuba	CSP	Cuba	CSP	Cuba
1995	-	2	-	14	-	195	-	5
1996	-	2	-	16	-	205	-	5
1997	-	1	-	9	-	117	-	12
1998	-	2	-	16	-	160	-	11
1999	-	2	-	18	-	237	-	15
2000	-	2	-	19	-	175	-	14
2001	-	2	-	16	-	241	-	15
2002	-	3	-	24	-	390	-	38
2003	-	4	-	46	-	954	-	18
2004	-	4	-	42	-	585	-	36
2005	-	4	-	34	-	695	-	15
2006	-	4	-	56	-	548	-	46
2007	4	4	55	40	1.269	868	29	77
2008	4	4	53	59	921	871	39	118
2009	4	4	51	51	1.105	954	51	185
2010	4	4	56	56	1.278	1.278	59	239
2011	4	5	55	55	1.067	1.456	73	327
2012	5	5	64	64	1.507	1.507	96	325
2013	-	2	-	31	-	843	-	109
	25	60	334	666	7.147	12.279	347	1.610

Fonte: desenvolvida pelo autor

Para o período entre 1995 e 2006, e 2013, somente existem informações da Coleção de Cuba. Todos os anos apresentaram diferenças nas informações sobre as citações recebidas, e somente os anos de 2010 e 2012 tiveram os mesmos totais de citações concedidas. Os anos de 2007 e 2008 apresentaram diferenças também em relação ao total de artigos. O ano de 2011, apesar de apresentar os mesmos totais de artigos, mostrou uma diferença no número de fascículos.

O ISSN 1135-5727 apresentou todas as informações compatíveis entre as Coleções da Espanha e de Saúde Pública para os anos de 1997 e 1998. Para os outros anos entre 1999 e 2013, foram encontradas diferenças no número de citações recebidas:

TABELA 6 – Dados fontes do SciELO: diferenças para o ISSN 1135-5727

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	Espanha	CSP	Espanha	CSP	Espanha	CSP	Espanha
1997	6	6	36	36	1.151	1.151	2	2
1998	6	6	51	51	1.203	1.203	15	15
1999	6	6	53	53	1.262	1.262	52	51
2000	6	6	54	54	1.447	1.447	22	24
2001	6	6	41	41	1.118	1.118	28	38
2002	6	6	56	56	1.712	1.712	47	52
2003	6	6	46	46	1.395	1.415	54	83
2004	6	6	47	47	1.455	1.455	33	66
2005	6	6	49	49	1.545	1.545	46	91
2006	6	6	46	46	46	1.400	18	109
2007	6	6	46	46	1.658	1.557	78	136
2008	6	6	49	49	1.762	1.747	115	207
2009	6	6	66	66	2.594	2.594	149	270
2010	6	6	60	60	1.749	1.749	127	278
2011	6	6	55	55	1.757	1.757	92	294
2012	6	6	50	50	1.843	1.843	93	270
2013	-	1	-	6	-	280	-	54
	96	97	805	811	23.697	25.235	971	2.040

Fonte: desenvolvida pelo autor

Além disso, diferenças no total de citações concedidas foram identificadas para os anos de 2003, 2006, 2007 e 2008. O ano de 2013 apresentou os dados somente da Coleção da Espanha.

Todas as informações de citações recebidas do ISSN 1413-8123 identificadas apresentaram divergência para o período 1998 a 2013:

TABELA 7 – Dados fontes do SciELO: diferenças para o ISSN 1413-8123

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	Brasil	CSP	Brasil	CSP	Brasil	CSP	Brasil
1998	1	1	13	13	210	210	5	12
1999	2	2	30	30	1.060	1.060	9	11
2000	2	2	27	27	1.063	1.063	9	17
2001	2	2	33	33	793	793	31	45
2002	4	4	64	64	1.994	1.994	51	64
2003	4	4	76	76	2.484	2.484	78	107
2004	4	4	85	85	2.411	2.411	120	175
2005	5	5	118	118	3.560	3.560	192	299
2006	5	5	109	109	2.907	2.907	135	284
2007	7	7	162	162	4.856	4.856	179	428
2008	8	8	211	211	6.514	6.514	352	761
2009	7	7	212	212	6.494	6.494	293	1.034
2010	9	9	356	356	10.350	10.350	278	1.139
2011	13	13	457	457	14.609	14.609	376	1.412
2012	12	12	308	308	10.311	10.311	412	1.666
2013	5	6	144	174	4.622	5.661	98	497
	90	91	2.405	2.435	74.338	75.377	2.618	7.951

Fonte: desenvolvida pelo autor

Além disso, o ano de 2013 apresentou diferenças em todas as informações.

Considerado o período entre 1985 e 2013, o ISSN 0102-311X apresentou diferença nos totais artigos e citações recebidas para o ano de 1987. Os anos de 1985, 1986 e 1988 não apresentaram nenhuma divergência. As citações recebidas divergem entre as duas coleções analisadas em todo o período entre 1989 e 2013. Nos anos de 2003 e 2007 também há diferença no total de artigos e citações concedidas.

TABELA 8 – Dados fontes do SciELO: diferenças para o ISSN 0102-311X

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	Brasil	CSP	Brasil	CSP	Brasil	CSP	Brasil
1985	4	4	25	25	407	407	3	3
1986	4	4	20	20	377	377	7	7
1987	4	4	25	24	445	445	2	3
1988	4	4	17	17	407	407	5	5
1989	4	4	18	18	269	269	6	9
1990	4	4	29	29	558	558	17	18
1991	4	4	38	38	869	869	20	22
1992	4	4	46	46	887	887	16	19
1993	5	5	56	56	1.336	1.336	29	31
1994	6	6	73	73	1.857	1.857	31	35
1995	4	4	54	54	1.253	1.253	41	49
1996	6	6	77	77	1.778	1.778	74	96
1997	6	6	96	96	2.312	2.312	76	100
1998	7	7	131	131	3.082	3.082	136	178
1999	6	6	110	110	3.083	3.083	143	192
2000	6	6	131	131	3.776	3.776	154	214
2001	7	7	158	158	4.765	4.765	203	335
2002	7	7	189	189	5.728	5.728	298	443
2003	8	8	230	229	6.077	6.013	372	587
2004	8	8	215	215	6.180	6.180	487	808
2005	7	7	214	214	5.923	5.923	730	1.244
2006	13	13	260	260	7.845	7.845	892	1.713
2007	16	16	337	338	10.365	10.405	1.088	2.144
2008	16	16	347	347	10.974	10.974	1.500	2.978
2009	15	15	301	301	9.945	9.945	1.468	3.314
2010	12	12	214	214	6.872	6.872	1.510	3.723
2011	15	15	264	264	8.752	8.752	1.955	4.194
2012	13	13	228	228	7.423	7.423	1.700	4.238
2013	5	5	88	88	3.158	3.158	618	1.348
	220	220	3.991	3.990	116.703	116.679	13.581	28.050

Fonte: desenvolvida pelo autor

O oitavo ISSN analisado, 0034-8910, mostrou informações compatíveis para o período 1967 a 1979 e para os anos 1981 e 1982. Todos os outros anos mostraram diferenças no total de citações recebidas. O ano de 2012 foi o único a apresentar divergência do total de citações concedidas para todo o período identificado: 1967 a 2013. Esses resultados são apresentados na TAB. 9 a seguir:

TABELA 9 – Dados fontes do SciELO: diferenças para o ISSN 0034-8910

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	Brasil	CSP	Brasil	CSP	Brasil	CSP	Brasil
1967	2	2	21	21	201	201	1	1
1968	2	2	13	13	261	261	0	0
1969	2	2	26	26	199	199	7	7
1970	2	2	24	24	178	178	17	17
1971	2	2	28	28	294	294	15	15
1972	4	4	43	43	310	310	14	14
1973	4	4	41	41	435	435	37	37
1974	5	5	45	45	513	513	18	18
1975	4	4	55	55	777	777	53	53
1976	6	6	40	40	605	605	24	24
1977	4	4	54	54	734	734	50	50
1978	4	4	51	51	836	836	60	60
1979	6	6	44	44	688	688	45	45
1980	4	4	52	52	1.108	1.108	83	84
1981	7	7	64	64	1.239	1.239	85	85
1982	7	7	34	34	664	664	60	60
1983	6	6	40	40	1.084	1.084	93	94
1984	7	7	53	53	1.020	1.020	67	73
1985	6	6	56	56	976	976	68	81
1986	6	6	53	53	865	865	64	81
1987	6	6	51	51	1.100	1.100	80	89
1988	6	6	61	61	1.311	1.311	91	117
1989	6	6	62	62	1.187	1.187	88	116
1990	7	7	67	67	1.629	1.629	91	120
1991	6	6	68	68	1.578	1.578	89	112
1992	6	6	64	64	1.346	1.346	84	122
1993	7	7	65	65	1.592	1.592	137	162
1994	6	6	65	65	1.599	1.599	121	167
1995	6	6	69	69	1.573	1.573	143	185
1996	6	6	78	78	1.929	1.929	163	256
1997	7	7	88	88	2.080	2.080	231	303
1998	6	6	85	85	1.883	1.883	170	306
1999	6	6	79	79	1.479	1.479	166	314
2000	7	7	112	112	1.874	1.874	261	430
2001	6	6	87	87	1.325	1.325	250	517
2002	7	7	121	121	2.360	2.360	246	586
2003	6	6	115	115	2.072	2.072	369	710
2004	7	7	129	129	2.254	2.254	465	921
2005	6	6	131	131	2.378	2.378	518	1.108
2006	8	8	161	161	3.285	3.285	564	1.392
2007	8	8	157	157	3.048	3.048	788	1.781
2008	8	8	170	170	3.257	3.257	1.063	2.361
2009	8	8	157	157	3.056	3.056	945	2.355
2010	6	6	132	132	2.764	2.764	975	2.689
2011	6	6	137	137	3.168	3.168	1.262	3.064
2012	7	7	138	138	3.039	3.043	1.148	3.147
2013	2	2	25	25	567	567	387	963
	263	263	3.511	3.511	67.720	67.724	11.756	25.292

Fonte: desenvolvida pelo autor

O ISSN 1415-790x, último da relação identificada, apresentou todos os dados idênticos para o primeiro ano disponível: 1998. Todos os anos entre 1999 e 2013 apresentaram diferença do total de citações recebidas:

TABELA 10 – Dados fontes do SciELO: diferenças para o ISSN 1415-790x

ANO	FASCÍCULOS		ARTIGOS		CITAÇÕES CONCEDIDAS		CITAÇÕES RECEBIDAS	
	CSP	Brasil	CSP	Brasil	CSP	Brasil	CSP	Brasil
1998	3	3	22	22	787	787	1	1
1999	2	2	9	9	208	208	3	4
2000	1	1	5	5	161	161	7	9
2001	3	3	15	15	521	521	12	13
2002	4	4	33	33	951	951	9	13
2003	4	4	33	33	706	706	17	23
2004	4	4	38	38	915	915	24	38
2005	5	5	42	42	1.016	1.016	27	41
2006	4	4	41	41	1.168	1.168	59	97
2007	4	4	58	58	1.568	1.568	100	167
2008	5	5	73	73	2.297	2.297	147	277
2009	4	4	61	61	1.875	1.875	160	357
2010	4	4	62	62	2.025	2.025	206	485
2011	5	5	80	80	2.300	2.300	261	569
2012	4	4	79	79	2.499	2.499	259	598
2013	1	1	21	21	603	603	100	169
	57	57	672	672	19.600	19.600	1.392	2.861

Fonte: desenvolvida pelo autor

4.3 Vinculação às áreas de conhecimento

Dos 904 periódicos avaliados das 10 coleções, 817 estavam vinculados somente a uma área de conhecimento, 73 a duas áreas, 9 a 3 áreas (0041-4751, 0185-3309, 0186,2979, 0253-570x, 0258-6576, 1315-0138, 1316-4821, 1690-4648 e 2305-7963), 2 a 4 áreas (0378-1844 e 0798-2259), 2 a 5 áreas (0038-2353) e 1 a 6 áreas (0001-3765).

Os 1.014 periódicos – consideradas as repetições nas áreas – foram agrupados de acordo com as áreas de conhecimento identificadas: Ciências da Saúde (290; 28,6%), Ciências Humanas (226; 22,3%), Ciências Sociais Aplicadas (148; 14,6%), Ciências Agrárias (96; 9,5%), Ciências Biológicas (94; 9,3%), Engenharias (70; 6,9%), Ciências Exatas e da Terra (67; 6,6%) e Linguística, Letras e Artes (23; 2,3%). Os totais de fascículos, artigos, citações concedidas e recebidas, bem como o total de periódicos correntes e não correntes são apresentados a seguir para cada área de conhecimento identificada:

FIGURA 20 – Periódicos SciELO agrupados por área de conhecimento

Áreas de Conhecimento

	Periódicos	Fascículos		Artigos		Citações Concedidas		Citações Recebidas		Correntes	Não correntes
8 áreas	904	29.728	100,00%	373.370	100,00%	9.823.290	100,00%	729.546	100,00%	781	123
CIENCIAS AGRARIAS	96	3.133	10,54%	58.057	15,55%	1.383.257	14,08%	162.645	22,29%	80	16
CIENCIAS BIOLOGICAS	94	3.472	11,68%	49.493	13,26%	1.382.040	14,07%	102.499	14,05%	83	11
CIENCIAS DA SAUDE	290	12.565	42,27%	151.513	40,58%	3.769.443	38,37%	333.465	45,71%	238	52
CIENCIAS EXATAS E DA TERRA	67	2.052	6,90%	28.331	7,59%	818.982	8,34%	31.946	4,38%	49	18
CIENCIAS HUMANAS	226	4.232	14,24%	44.425	11,90%	1.332.891	13,57%	61.560	8,44%	206	20
CIENCIAS SOCIAIS APLICADAS	148	2.405	8,09%	20.573	5,51%	633.131	6,45%	23.364	3,20%	136	12
ENGENHARIAS	70	1.545	5,20%	18.208	4,88%	427.500	4,35%	12.677	1,74%	61	9
LINGUISTICA LETRAS E ARTES	23	324	1,09%	2.770	0,74%	76.046	0,77%	1.390	0,19%	23	0

Fonte: desenvolvida pelo autor¹²

¹² Disponível em: <http://www.cmca.srv.br/prototipo/metabuscaador_index_area.php?ss=22357&id=dfb293d561eb534aa9ff2c18f6045eaa>. Acesso em: 13 jun. 2013.

5 RESULTADOS DA FASE II: BASE DE CITAÇÕES DA PERSPECTIVAS EM CIÊNCIA DA INFORMAÇÃO

O teste do protótipo, cujos resultados são apresentados neste capítulo, avaliou a criação automática da base de citações para o periódico PCI, representante do estrato Qualis A1 da área de Sociais Aplicadas I da CAPES e única revista da área de CI indexada no SciELO à época. Foram usados os artigos do periódico PCI – publicados entre 2006 e 2012, até o volume 3, e o volume 1 do ano de 2013 – disponíveis na data de realização dos testes.

Como o ponto fulcral desta pesquisa é a proposta de uma metodologia, de um processo, não se pretende avaliar ou corrigir o conteúdo dos arquivos XML importados, nem realizar análises qualitativas das informações apresentadas. Após a criação da base de citações, seu uso será exemplificado produzindo-se relatórios no mesmo formato daqueles criados por Garfield (FIG. 6, FIG. 7 e FIG. 8), embora outros possam ser gerados a partir dos dados armazenados.

5.1 Obtenção dos dados do SciELO e criação da Base de Citações

O SciELO apresenta um resumo estatístico para os periódicos indexados, denominado “Lista de dados fonte”, e no caso da PCI a FIG. 21 apresenta os seguintes dados:

FIGURA 21 – Lista de dados fonte: Perspectivas em Ciência da Informação



Perspectivas em Ciência da Informação
ISSN 1413-9936

Data do último processamento
24-03-2013

Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

título da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Perspect. ciênc. inf	24	290	7714	231	12.08
2013	1	10	289	4	10.00
2012	3	34	1038	29	11.33
2011	4	50	1365	45	12.50
2010	3	39	1192	51	13.00
2009	4	55	1353	33	13.75
2008	3	39	1050	31	13.00
2007	3	35	873	17	11.67
2006	3	28	554	21	9.33
total	24	290	7714	231	12.08

Fonte: SciELO, 2013¹³

De acordo com esse relatório, para o período definido entre 2006 e 2013, estão disponíveis 24 fascículos, 290 artigos e 7.714 citações para a PCI. Esses dados foram usados para comparar os resultados obtidos no protótipo, que são detalhados a seguir:

TABELA 11 – Resultados da importação XML: PCI 2006 a 2013

ANO	EDIÇÕES	ARTIGOS	MÉDIA DE ARTIGOS	CITAÇÕES
2013	1	11	11,00	289
2012	3	34	11,33	1038
2011	4	50	12,50	1365
2010	3	40	13,33	1192
2009	4	55	13,75	1353
2008	3	39	13,00	1050
2007	3	35	11,67	873
2006	3	28	9,33	554
TOTAL	24	292	12,17	7714

Fonte: desenvolvida pelo autor

¹³ Disponível em: <
[http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=scl&issn=1413-9936&CITED\[\]=PERSPECTIVAS+EM+CIENCIA+DA+INFORMACAO&YNG\[\]=all](http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=scl&issn=1413-9936&CITED[]=PERSPECTIVAS+EM+CIENCIA+DA+INFORMACAO&YNG[]=all)>. Acesso em: 05 abr. 2013.

Assim como os dados do SciELO, o protótipo também apresentou 24 fascículos (edições) e 7.714 citações. Entretanto, foram encontrados 2 artigos a mais do que o número encontrado para o SciELO – considerando apenas artigos que possuem ao menos uma citação. Uma das diferenças observadas ocorreu no total de artigos para o ano de 2010, com 1 artigo a menos no relatório do SciELO. Os arquivos XML foram conferidos para este ano, e nova pesquisa manual ao SciELO mostrou que uma resenha¹⁴ apresenta exatamente 7 referências – entretanto, esta resenha parece não estar incluída no total de 39 artigos do ano de 2010 de acordo com os dados do SciELO (FIG. 21).

Cada um dos 292 arquivos XML da PCI foi interpretado e as informações foram armazenadas no Módulo “Base de Citações”. A comparação anual entre os dados obtidos automaticamente do SciELO e as informações armazenadas a partir da interpretação dos arquivos XML é apresentada a seguir.

O resumo da importação apresenta, para cada ano, o total de fascículos, de artigos, a média (de artigos em cada fascículo) e o total de citações. Essas informações são apresentadas para a base de citações (BC) e para o SciELO, tendo como fontes, respectivamente, a interpretação dos arquivos XML e a lista de dados fonte do periódico.

Eventuais diferenças entre os números da BC e do SciELO são identificadas nas cores azul (quando o número da BC é superior ao do SciELO) e vermelho (se o número do SciELO for superior ao da BC). Quando há divergência entre os números da BC e do SciELO, é apresentado o percentual em relação ao número do SciELO. Caso os números da BC e do SciELO sejam idênticos, são apresentados na cor verde.

A seguir são apresentadas as FIGURAS correspondentes respectivamente ao resumo da importação de dados do SciELO (FIG. 22), aos artigos do ano de 2013 para o periódico PCI (FIG. 23), e às referências do primeiro artigo listado (FIG. 24):

¹⁴ BOCCATO, Vera Regina Casari. **Perspect. ciênc. inf.**, Belo Horizonte, v. 15, n. 3, 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362010000300014&lng=pt&nrm=iso>. Acesso em: 01 jan. 2013.

FIGURA 22 – Resumo da importação de dados do SciELO: dados fonte e arquivos XML

1413-9936	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	24	24	0	292	290	2 (0,69%)	12,17	12,08	0,09 (0,74%)	7.714	7.714	0
2013	1	1	0	11	10	1 (10,00%)	11,00	10,00	1,00 (10,00%)	289	289	0
2012	3	3	0	34	34	0	11,33	11,33	0,00	1.038	1.038	0
2011	4	4	0	50	50	0	12,50	12,50	0,00	1.365	1.365	0
2010	3	3	0	40	39	1 (2,56%)	13,33	13,00	0,33 (2,54%)	1.192	1.192	0
2009	4	4	0	55	55	0	13,75	13,75	0,00	1.353	1.353	0
2008	3	3	0	39	39	0	13,00	13,00	0,00	1.050	1.050	0
2007	3	3	0	35	35	0	11,67	11,67	0,00	873	873	0
2006	3	3	0	28	28	0	9,33	9,33	0,00	554	554	0

Fonte: desenvolvida pelo autor¹⁵

¹⁵ Disponível em: < http://cmca.srv.br/prototipo/metabuscador_mostraisn.php?issn=1413-9936>. Acesso em: 05 abr. 2013.

FIGURA 23 – Artigos listados do ISSN 1413-9936 para o ano de 2013

Artigos: 11

#	Autoria	Ano	Periódico	Título 	Volume	Número	Páginas	Referências			
1	MARQUES, Rodrigo Moreno PERELMAN, Michael	2013	Perspect. ciênc. inf.		18	1	179-185	6			
2	FERREIRA, Leticia Elaine ALMEIDA JÚNIOR, Oswaldo Francisco de	2013	Perspect. ciênc. inf.	A mediação da informação no âmbito da arquivística	18	1	158-167	12			
3	MARAVILHAS-LOPES, Sérgio Paulo	2013	Perspect. ciênc. inf.	A Web 2.0 como ferramenta de análise de tendências e monitorização do ambiente externo e sua relação com a cultura de convergência dos media	18	1	126-137	58			
4	ROSA, Juan Miguel VERAS, Manoel	2013	Perspect. ciênc. inf.	Avaliação heurística de usabilidade em jornais online: estudo de caso em dois sites	18	1	138-157	13			
5	MACIEL FILHO, Adalberto Rego AQUINO, Miriam Cunha SOARES, Islla Silva SILVA, Camila Alves	2013	Perspect. ciênc. inf.	Bibliotecas domiciliares: um estudo de caso	18	1	26-42	17			
6	BASTOS, Gustavo Grandini GALLI, Fernanda Correa Silveira ROMÃO, Lucília Maria Sousa	2013	Perspect. ciênc. inf.	Discursividades sobre o bibliotecário	18	1	2-14	20			
7	FIALHO, Janaina	2013	Perspect. ciênc. inf.	Experiência com estudantes do ensino médio através da pesquisa escolar orientada	18	1	15-25	6			
8	TAROUCO, Liane Margarida Rockenbach RODRIGUES, Alessandra Pereira SCHMITT, Marcelo Augusto Rauh TODESCO, José Leomar	2013	Perspect. ciênc. inf.	Integração do MOODLE com repositórios abertos	18	1	66-85	35			
9	STEIL, Andrea Valeria ZANCANARO, Airton ERPEN, Julio Graeff SANTOS, Jane Lucia S.	2013	Perspect. ciênc. inf.	Mapeamento da produção científica sobre memória organizacional e ontologias	18	1	43-65	66			
10	SOUZA, Renato Rocha TEIXEIRA, Renata Cristina OLIVEIRA, Paulo Henrique de GONÇALVES, Carlos Alberto	2013	Perspect. ciênc. inf.	O uso das informações contidas em documentos de patentes nas práticas de Inteligência Competitiva: apresentação de um estudo das patentes da UFMG	18	1	106-125	18			
11	DE PAULA, Edmar Aderson Mendes ASSIS, Vivian Coelho	2013	Perspect. ciênc. inf.	Um estudo sobre gestão do conhecimento e inovação em uma empresa multinacional do setor de fast-food: o caso da Subway	18	1	86-105	38			

Fonte: desenvolvida pelo autor¹⁶¹⁶ Disponível em: < http://cmca.srv.br/prototipo/metabuscaador_mostraiissn.php?issn=1413-9936>. Acesso em: 05 abr. 2013.

FIGURA 24 – Referências citadas em um artigo de 2013 do ISSN 1413-9936

Citações: 6

#	Autoria	Ano	Título	Fonte	Tipo	Citado por	Ano	Periódico	Artigo 
1	PERELMAN, Michael.	1991		Information, Social Relations, and the Economics of High Technology	book	MARQUES, Rodrigo Moreno PERELMAN, Michael	2013	Perspect. ciênc. inf.	  
2	PERELMAN, Michael.	1998		Class Warfare in the Information Age	book	MARQUES, Rodrigo Moreno PERELMAN, Michael	2013	Perspect. ciênc. inf.	  
3	PERELMAN, Michael.	2002		Steal this idea	book	MARQUES, Rodrigo Moreno PERELMAN, Michael	2013	Perspect. ciênc. inf.	  
4	_____	Janu	The Political Economy of Intellectual Property	Montly Review	journal	MARQUES, Rodrigo Moreno PERELMAN, Michael	2013	Perspect. ciênc. inf.	  
5	_____	2003	Intellectual Property Rights and the Commodity Form: New Dimensions in the Legislated Transfer of Surplus Value	Review of Radical Political Economics	journal	MARQUES, Rodrigo Moreno PERELMAN, Michael	2013	Perspect. ciênc. inf.	  
6	_____	Febr	What Went Wrong: An Idiosyncratic Perspective on the Economy and Economics	Review of Radical Political Economics	journal	MARQUES, Rodrigo Moreno PERELMAN, Michael	2013	Perspect. ciênc. inf.	  

Fonte: desenvolvida pelo autor¹⁷

¹⁷ Disponível em: < http://cmca.srv.br/prototipo/metabuscaador_mostraiissn.php?issn=1413-9936>. Acesso em: 05 abr. 2013.

Na FIG. 24 observa-se a ocorrência de um dos problemas citados anteriormente em relação às datas alfanuméricas para as citações 4 e 6.

Para auxiliar na validação das informações, o programa desenvolvido também apresenta um resumo das estruturas de *tags* identificadas para o periódico:

FIGURA 25 – Estrutura das *tags* XML dos arquivos interpretados

ESTRUTURA DOS ARTIGOS NA BC (<i>tags</i> XML) - 1413-9936	ARTIGOS BC	CITAÇÕES BC
article; front; /front; back; ref-list; ref-id; /ref; /ref-list; /back; /article;	292	7.714
TOTAL	292	7.714

Fonte: desenvolvida pelo autor¹⁸

A FIG. 25 permite identificar 292 artigos incorporados à base de citações que possuem, no total, 7.714 citações associadas. O total de artigos apresentado é idêntico ao total de artigos da BC apresentado na FIG. 22, bem como o total de citações.

Além desses dados, também é apresentado um resumo dos arquivos incorporados à BC, porém sem citações associadas:

FIGURA 26 – Arquivos XML incorporados sem citações associadas

SITUAÇÃO - 1413-9936	ESTRUTURA DOS ARTIGOS NA BC (<i>tags</i> XML)	ARTIGOS BC	CITAÇÕES BC
Artigos sem citações incorporadas	article; front; /front; /article; Referências identificadas: 0/0 Ano: 2013	8	0

Fonte: desenvolvida pelo autor¹⁹

Para o caso da PCI, foram listados os 8 artigos nessa situação (FIG. 27). Percebe-se, pela estrutura de *tags* apresentada, que não existe a identificação das referências nestes artigos (todos eles pertencentes ao ano de 2013): *tags back, ref-list, ref-id, /ref-id, /ref-list e /back*. Finalizando a validação da importação, cada um desses 8 artigos pode ser acessado para confirmação, sendo possível inclusive o acesso ao resumo e ao texto completo na base SciELO:

¹⁸ Disponível em: < http://cmca.srv.br/prototipo/metabuscaador_mostraisn.php?issn=1413-9936>. Acesso em: 05 abr. 2013.

¹⁹ Disponível em: < http://cmca.srv.br/prototipo/metabuscaador_mostraisn.php?issn=1413-9936>. Acesso em: 05 abr. 2013.

FIGURA 27 – Relação de artigos sem citações associadas, na PCI, para o ano de 2013

Artigos: 8

#	Autoria	Ano	Periódico	Título 	Volume	Número	Páginas	Referências			
1	OLIVEIRA, Marlene de NEVES, Jorge Tadeu de Ramos	2013	Perspect. ciênc. inf.		18	1	1-1	0			
2	SEPÚLVEDA, Maria Inês Moreira	2013	Perspect. ciênc. inf.	A relação dos bibliotecários com a profissão, com a rotina profissional e com os usuários a partir de uma perspectiva compreensiva	18	1	178-178	0			
3	SANTOS, Vânia Carvalho Rôla	2013	Perspect. ciênc. inf.	Gestão, informação e comunicação museológica: um estudo comparativo entre pequenos e médios museus brasileiros e franceses	18	1	171-172	0			
4	GANDRA, Tatiane Krempser	2013	Perspect. ciênc. inf.	Inclusão digital na terceira idade: um estudo de usuários sob a perspectiva fenomenológica	18	1	175-176	0			
5	SANTOS, Patricia Espírito	2013	Perspect. ciênc. inf.	O que informam as cartas de leitores e leitoras enviadas a jornais impressos: os casos do Estado de Minas e do Le Monde	18	1	170-170	0			
6	MACEDO, Solange Madalena Souza	2013	Perspect. ciênc. inf.	Orientação informacional em instituições de ensino superior de Belo Horizonte	18	1	177-177	0			
7	FERNANDES, Wesley Rodrigo	2013	Perspect. ciênc. inf.	Portal Periódicos CAPES: estudo dos não- usuários docentes das IFES brasileiras	18	1	173-174	0			
8	CAVALCANTE, Ricardo Bezerra	2013	Perspect. ciênc. inf.	Sistema de Informação da Atenção Básica (SIAB) como instrumento de poder	18	1	168-169	0			

Fonte: desenvolvida pelo autor²⁰

²⁰ Disponível em:

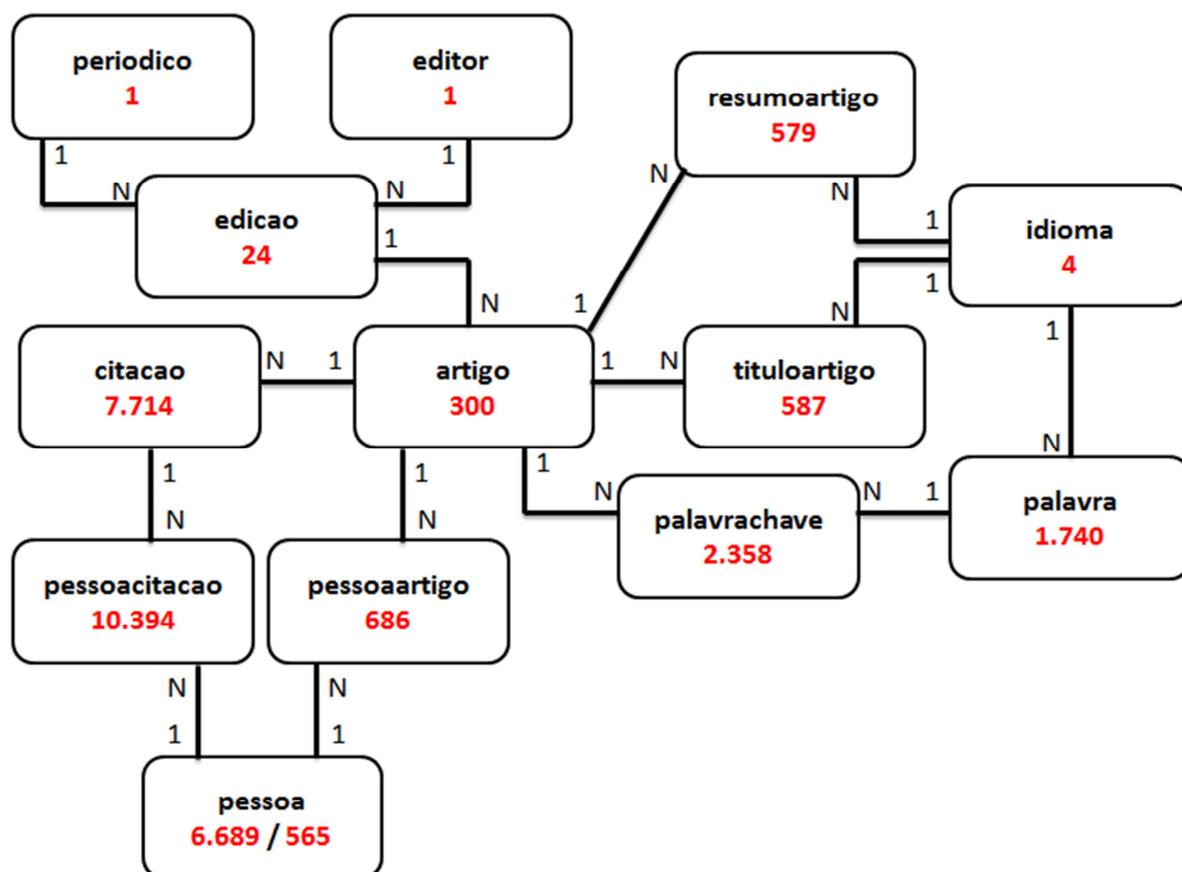
<http://www.cmca.srv.br/prototipo/metabuscador_mostraartigo.php?art=23917,23904,23903,23901,23920,23911,23914,23918&ss=339172512&id=5e8eaacaa6791d624b67f40de34f1456>. Acesso em: 05 abr. 2013.

5.2 Descrição da amostra: a PCI em números

Para exemplificar o uso da base de citações, a seguir são apresentados dados quantitativos gerados a partir da importação automática das informações do SciELO conforme a metodologia proposta.

Em termos de quantidade de registros, a PCI apresentou os seguintes valores para cada tabela do banco de dados, em relação aos periódicos citantes: 1 periódico (PCI), 1 editor (Escola de Ciência da Informação da UFMG), 24 fascículos (tabela *edicao*), 300 artigos (292 apresentaram citações associadas a eles; 8 não) e 7.714 citações. Foram encontrados 579 resumos (tabela *resumoartigo*) – 286 em português, 283 em inglês, 5 em espanhol e 5 em francês – e 587 títulos (tabela *tituloartigo*), dos quais 293 em português, 283 em inglês, 6 em espanhol e 5 em francês. A FIG. 28 apresenta os totais de registros em cada tabela:

FIGURA 28 – Quantidade de registros incorporados no banco de dados: PCI



Fonte: desenvolvida pelo autor²¹

²¹ Consulta ao banco de dados realizada em 07 abr. 2013.

Do total de 2.358 palavras-chave (tabela *palavrachave*) apresentadas em todos os artigos, 1.740 correspondem a termos distintos (tabela *palavra*) que foram apresentados em inglês (851), em português (835), em francês (28) e em espanhol (26). As 10 palavras que mais ocorreram nesta amostra foram: Ciência da Informação (25), *Information Science* (22), Gestão do Conhecimento e *Knowledge Management* (19 cada), Bibliometria (15), Informação e *Information* (14 cada), *Internet* e Produção Científica (13 cada) e *Scientific Production* (12).

Foram identificados 686 autores de artigos (tabela *pessoaartigo*) sendo 565 nomes distintos (tabela *pessoa*). Os 10 autores com mais artigos produzidos – ordenados pela quantidade decrescente de artigos e ordem alfabética do sobrenome – foram: Bufrem, Leilah Santiago (6 artigos), Araujo, Carlos Alberto Ávila; Boccato, Vera Regina Casari; Cunha, Murilo Bastos da; Fujita, Mariângela Spotti Lopes; Nagano, Marcelo Seido; Souza, Renato Rocha; Todesco, José Leomar (cada um com 4 artigos) e ARAÚJO, Eliany Alvarenga de e Arboit, Aline Elis (cada um com 3 artigos²²).

Em relação aos periódicos citados, dos 10.394 autores identificados nas citações (tabela *pessoacitacao*), 6.689 são distintos (tabela *pessoa*). Os 10 nomes mais citados foram: Nonaka, I. (69 ocorrências), Choo, C. W. (35), Pinheiro, L. V. R. (33), Castells, M. (32), Davenport, T. H. e Takeuchi, H. (29 cada), Capurro, R. e González de Gómez (28 cada), Saracevic, T. (27) e Lévy, P. (26).

As fontes mais citadas, desconsiderados os valores “branco” (116 ocorrências), “Anais...” (209), “Anais” (56) e “Proceedings...” (30), e sem nenhum tipo de recorte, foram: Ciência da Informação (273), Perspectivas em Ciência da Informação (108), *Scientometrics* (71), Transinformação (55), *Journal of Documentation* (47), DataGramZero (45), *Journal of the American Society for Information Science* (45), *Journal of Information Science* (44), *Journal of knowledge Management* (35) e *Ci. Inf.* (28).

As próximas FIGURAS deste tópico descrevem a amostra nos formatos dos relatórios apresentados por Garfield (1972, p.527-30): frequências de citações (FIG. 29), estatísticas dos periódicos citados (FIG. 30) e estatísticas dos periódicos citantes (FIG. 31). O primeiro relatório acumula o número de vezes que uma referência foi citada, e distribui essas citações por ano em que foram citadas. Foram considerados os 10 periódicos mais citados listados anteriormente e detalhados os 10 últimos anos, com o total dos anos anteriores acumulados na última coluna. O segundo, similar ao primeiro, detalha para cada fonte citada os periódicos citantes. No caso específico de um único ISSN, o detalhamento

²² Outros autores também apresentaram 3 artigos produzidos, entretanto foram listados no total 10 autores, em ordem alfabética.

não apresenta diferença alguma do total. A terceira e última lista produzida é similar à segunda, entretanto organiza os dados por periódico citante, detalhando os periódicos citados. Novamente, para um único periódico, somente existe um periódico citante.

Para todos os relatórios, foram considerados os 10 periódicos mais citados listados anteriormente e detalhados os 10 últimos anos, com o total dos anos anteriores acumulados na última coluna.

FIGURA 29 – Frequências de citações: PCI

FONTE CITADA	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Ciência da Informação	273	2	42	21	38	52	45	40	33	-	-	-
Perspectivas em Ciência da Informação	108	2	10	21	30	13	19	7	6	-	-	-
Scientometrics	71	-	15	6	13	-	23	9	5	-	-	-
Transinformação	55	2	11	11	12	5	3	4	7	-	-	-
Journal of Documentation	47	-	5	8	12	5	8	7	2	-	-	-
DataGramaZero	45	3	16	8	3	7	4	1	3	-	-	-
Journal of the American Society for Information Science	45	-	4	6	3	3	6	16	7	-	-	-
Journal of Information Science	44	-	4	5	12	8	6	4	5	-	-	-
Journal of knowledge Management	35	2	-	10	16	4	1	2	-	-	-	-
Ci. Inf.	28	1	7	10	-	5	2	3	-	-	-	-

Fonte: desenvolvida pelo autor²³

²³ Consulta ao banco de dados realizada em 10 abr. 2013.

FIGURA 30 – Estatística dos periódicos citados: PCI

FONTE CITADA	CITANTE	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Ciência da Informação		273	2	42	21	38	52	45	40	33	-	-	0
	Perspect. ciênc. inf.	273	2	42	21	38	52	45	40	33	-	-	0
Perspectivas em Ciência da Informação		108	2	10	21	30	13	19	7	6	-	-	0
	Perspect. ciênc. inf.	108	2	10	21	30	13	19	7	6	-	-	0
Scientometrics		71	-	15	6	13	-	23	9	5	-	-	0
	Perspect. ciênc. inf.	71	-	15	6	13	-	23	9	5	-	-	0
Transinformação		55	2	11	11	12	5	3	4	7	-	-	0
	Perspect. ciênc. inf.	55	2	11	11	12	5	3	4	7	-	-	0
Journal of Documentation		47	-	5	8	12	5	8	7	2	-	-	0
	Perspect. ciênc. inf.	47	-	5	8	12	5	8	7	2	-	-	0
DataGramZero		45	3	16	8	3	7	4	1	3	-	-	0
	Perspect. ciênc. inf.	45	3	16	8	3	7	4	1	3	-	-	0
Journal of the American Society for Information Science		45	-	4	6	3	3	6	16	7	-	-	0
	Perspect. ciênc. inf.	45	-	4	6	3	3	6	16	7	-	-	0
Journal of Information Science		44	-	4	5	12	8	6	4	5	-	-	0
	Perspect. ciênc. inf.	44	-	4	5	12	8	6	4	5	-	-	0
Journal of Knowledge Management		35	2	-	10	16	4	1	2	-	-	-	0
	Perspect. ciênc. inf.	35	2	-	10	16	4	1	2	-	-	-	0
Ci. Inf.		28	1	7	10	-	5	2	3	-	-	-	0
	Perspect. ciênc. inf.	28	1	7	10	-	5	2	3	-	-	-	0

Fonte: desenvolvida pelo autor²⁴

²⁴ Consulta ao banco de dados realizada em 11 abr. 2013.

FIGURA 31 – Estatística dos periódicos citantes: PCI

CITANTE	FONTE CITADA	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Perspect. ciênc. inf.		7.714	289	1.038	1.365	1.192	1.353	1.050	873	554	-	-	0
	Ciência da Informação	273	2	42	21	38	52	45	40	33	-	-	0
	Anais...	209	16	24	41	31	31	15	7	44	-	-	0
		116	1	32	18	9	13	25	11	7	-	-	0
	Perspectivas em Ciência da Informação	108	2	10	21	30	13	19	7	6	-	-	0
	Scientometrics	71	-	15	6	13	-	23	9	5	-	-	0
	Anais	56	-	-	3	9	30	-	9	5	-	-	0
	Transinformação	55	2	11	11	12	5	3	4	7	-	-	0
	Journal of Documentation	47	-	5	8	12	5	8	7	2	-	-	0
	DataGramZero	45	3	16	8	3	7	4	1	3	-	-	0
	Journal of the American Society for Information Science	45	-	4	6	3	3	6	16	7	-	-	0
	Outros	6.689	263	879	1.222	1.032	1.194	902	762	435	-	-	0

Fonte: desenvolvida pelo autor²⁵

²⁵ Consulta ao banco de dados realizada em 10 abr. 2013.

De forma similar aos relatórios apresentados anteriormente, as FIGURAS a seguir apresentam os autores mais citados (FIG. 32), as palavras-chave mais usadas (FIG. 33) e os autores que mais produziram artigos no periódico (FIG. 34). Foram considerados os 10 autores mais citados, classificados em ordem alfabética.

FIGURA 32 – Autores mais citados: PCI

CITANTE	AUTOR CITADO	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Perspect. ciênc. inf. (1413-9936)		10.394	406	1.227	1.966	1.694	1.784	1.487	1.094	736	-	-	0
	NONAKA, I.	69	<u>1</u>	<u>3</u>	<u>25</u>	<u>10</u>	<u>8</u>	<u>13</u>	<u>7</u>	<u>2</u>	-	-	0
	CHOO, C. W.	35	<u>1</u>	<u>3</u>	<u>5</u>	<u>6</u>	<u>6</u>	<u>9</u>	<u>5</u>	-	-	-	0
	PINHEIRO, L. V. R.	33	<u>1</u>	<u>6</u>	<u>10</u>	<u>7</u>	<u>1</u>	<u>4</u>	<u>2</u>	<u>2</u>	-	-	0
	CASTELLS, M.	32	<u>1</u>	<u>2</u>	<u>8</u>	<u>5</u>	<u>3</u>	<u>6</u>	<u>6</u>	<u>1</u>	-	-	0
	DAVENPORT, T. H.	29	<u>1</u>	<u>4</u>	<u>3</u>	<u>4</u>	<u>7</u>	<u>2</u>	<u>6</u>	<u>2</u>	-	-	0
	TAKEUCHI, H.	29	<u>1</u>	<u>1</u>	<u>3</u>	<u>8</u>	<u>3</u>	<u>6</u>	<u>4</u>	<u>3</u>	-	-	0
	CAPURRO, R.	28	-	<u>4</u>	<u>5</u>	<u>4</u>	-	<u>5</u>	<u>10</u>	-	-	-	0
	GONZÁLEZ DE GÓMEZ, M. N.	28	-	<u>2</u>	<u>4</u>	<u>9</u>	<u>4</u>	<u>8</u>	-	<u>1</u>	-	-	0
	SARACEVIC, T.	27	-	<u>3</u>	<u>5</u>	<u>7</u>	<u>1</u>	<u>7</u>	<u>1</u>	<u>3</u>	-	-	0
	LÉVY, P.	26	<u>1</u>	<u>3</u>	<u>6</u>	<u>5</u>	<u>3</u>	<u>4</u>	<u>2</u>	<u>2</u>	-	-	0
	Outros	10.058	399	1.196	1.892	1.629	1.748	1.423	1.051	720	-	-	0

Fonte: desenvolvida pelo autor²⁶

²⁶ Consulta ao banco de dados realizada em 10 abr. 2013.

FIGURA 33 – Palavras-chave mais utilizadas: PCI

CITANTE	PALAVRA-CHAVE	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
	Perspect. ciênc. inf. (1413-9936)	2.358	78	287	396	335	452	332	259	219	-	-	0
	Ciência da informação (Português)	25	-	<u>3</u>	<u>5</u>	<u>4</u>	<u>4</u>	<u>3</u>	<u>2</u>	<u>4</u>	-	-	0
	Information Science (Inglês)	22	-	<u>3</u>	<u>5</u>	<u>2</u>	<u>4</u>	<u>2</u>	<u>2</u>	<u>4</u>	-	-	0
	Gestão do conhecimento (Português)	19	<u>1</u>	<u>1</u>	<u>1</u>	<u>4</u>	<u>4</u>	<u>2</u>	<u>4</u>	<u>2</u>	-	-	0
	Knowledge management (Inglês)	19	<u>1</u>	<u>1</u>	<u>2</u>	<u>4</u>	<u>3</u>	<u>2</u>	<u>4</u>	<u>2</u>	-	-	0
	Bibliometria (Português)	15	<u>1</u>	<u>3</u>	<u>1</u>	<u>2</u>	<u>1</u>	<u>2</u>	<u>4</u>	<u>1</u>	-	-	0
	Informação (Português)	14	-	-	<u>3</u>	-	<u>4</u>	<u>3</u>	<u>1</u>	<u>3</u>	-	-	0
	information (Inglês)	14	-	-	<u>3</u>	-	<u>4</u>	<u>3</u>	<u>1</u>	<u>3</u>	-	-	0
	Produção científica (Português)	13	<u>1</u>	<u>2</u>	<u>2</u>	<u>1</u>	<u>2</u>	<u>2</u>	<u>2</u>	<u>1</u>	-	-	0
	Scientific production (Inglês)	12	<u>1</u>	<u>2</u>	<u>2</u>	<u>1</u>	<u>2</u>	<u>2</u>	<u>2</u>	-	-	-	0
	Gestão da Informação (Português)	10	-	-	<u>1</u>	<u>2</u>	<u>1</u>	<u>2</u>	<u>4</u>	-	-	-	0
	Outros	2.195	73	272	371	315	423	309	233	199	-	-	0

Fonte: desenvolvida pelo autor²⁷

²⁷ Consulta ao banco de dados realizada em 10 abr. 2013.

FIGURA 34 – Autores que mais produziram: PCI

CITANTE	AUTOR PRODUTOR	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Perspect. ciênc. inf. (1413-9936)		686	38	72	124	89	114	94	79	76	-	-	0
	Bufrem, Leilah Santiago	6	-	-	<u>2</u>	<u>2</u>	<u>1</u>	-	<u>1</u>	-	-	-	0
	Araujo, Carlos Alberto Ávila	4	-	-	<u>1</u>	<u>1</u>	-	<u>1</u>	<u>1</u>	-	-	-	0
	Boccatto, Vera Regina Casari	4	-	-	-	<u>2</u>	<u>1</u>	-	-	<u>1</u>	-	-	0
	CUNHA, Murilo Bastos da	4	-	<u>1</u>	<u>1</u>	-	-	<u>1</u>	<u>1</u>	-	-	-	0
	Fujita, Mariângela Spotti Lopes	4	-	<u>1</u>	-	<u>1</u>	-	-	-	<u>2</u>	-	-	0
	Nagano, Marcelo Seido	4	-	<u>1</u>	<u>1</u>	-	<u>1</u>	<u>1</u>	-	-	-	-	0
	Souza, Renato Rocha	4	<u>1</u>	-	-	<u>1</u>	-	<u>1</u>	-	<u>1</u>	-	-	0
	Todesco, José Leomar	4	<u>1</u>	-	<u>1</u>	<u>2</u>	-	-	-	-	-	-	0
	ARAÚJO, Eliany Alvarenga de	3	-	-	<u>1</u>	-	-	-	<u>1</u>	<u>1</u>	-	-	0
	Arboit, Aline Elis	3	-	<u>1</u>	<u>1</u>	<u>1</u>	-	-	-	-	-	-	0
	Outros	646	36	68	116	79	111	90	75	71	-	-	0

Fonte: desenvolvida pelo autor²⁸

²⁸ Consulta ao banco de dados realizada em 10 abr. 2013.

6 RESULTADOS DA FASE II: BASE DE CITAÇÕES DA COLEÇÃO SAÚDE PÚBLICA

A validação final do protótipo, com uma amostra mais representativa e cujos resultados são apresentados a seguir, avaliou a criação automática da base de citações para a Coleção Saúde Pública do SciELO, utilizando todos os arquivos XML disponíveis à época.

Como o ponto fulcral desta pesquisa é a proposta de uma metodologia, de um processo, não se pretende avaliar ou corrigir o conteúdo dos arquivos XML importados, nem realizar análises qualitativas das informações apresentadas. Após a criação da base de citações, seu uso será exemplificado produzindo-se relatórios no mesmo formato daqueles criados por Garfield (FIG. 6, FIG. 7 e FIG. 8), embora outros possam ser gerados a partir dos dados armazenados.

6.1 Obtenção dos dados do SciELO e criação da Base de Citações

A lista de periódicos da Coleção Saúde Pública, obtida automaticamente do SciELO, apresentou 15 periódicos:

QUADRO 5 – Periódicos da Coleção Saúde Pública do SciELO

ISSN	Título
0021-2571	Annali dell'Istituto Superiore di Sanità
0034-8910	Revista de Saúde Pública
0036-3634	Salud Pública de México
0042-9686	Bulletin of the World Health Organization
0102-311X	Cadernos de Saúde Pública
0124-0064	Revista de Salud Pública
0213-9111	Gaceta Sanitaria
0864-3466	Revista Cubana de Salud Pública
1020-4989	Revista Panamericana de Salud Pública
1135-5727	Revista Española de Salud Pública
1413-8123	Ciência e Saúde Coletiva
1415-790x	Revista Brasileira de Epidemiologia
1555-7960	MEDICC Review
1726-4634	Revista Peruana de Medicina Experimental y Salud Pública
1851-8265	Salud Colectiva

Fonte: desenvolvido pelo autor

A “Lista de dados fonte” de cada um dos periódicos da Coleção Saúde Pública constantes do QUADRO 5 foi obtida automaticamente e gravada no banco de dados – a única exceção foi o periódico “*Annali dell'Istituto Superiore di Sanità*”, ISSN 0021-2571, pois

os dados não estavam disponíveis na data da consulta²⁹. A FIG. 35 apresenta um resumo para cada ISSN, respeitadas as mesmas considerações do capítulo anterior em relação às cores e percentuais calculados:

²⁹ Disponível em:
<http://statbiblio.scielo.org//stat_biblio/index.php?state=15&lang=pt&country=spa&issn=0021-2571&CITED%5B%5D=annali%20dellistituto%20superiore%20di%20sanita&YNG%5B%5D=all>. Acesso em: 01 abr. 2013.

FIGURA 35 – Resumo dos dados de importação das citações da Coleção de Saúde Pública do SciELO

ISSN	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA
TOTAL (14)	1.330	1.347	-17 (-1,26%)	18.689	17.672	1.017 (5,75%)	14,05	13,12	0,93 (7,09%)	491.613	488.252	3.361 (0,69%)
1851-8265	26	26	0	200	165	35 (21,21%)	7,69	6,35	1,34 (21,10%)	5.820	5.833	-13 (-0,22%)
1726-4634	12	12	0	311	278	33 (11,87%)	25,92	23,17	2,75 (11,87%)	6.509	6.509	0
1555-7960	6	6	0	55	46	9 (19,57%)	9,17	7,67	1,50 (19,56%)	1.554	1.560	-6 (-0,38%)
1415-790x	55	56	-1 (-1,79%)	666	651	15 (2,30%)	12,11	11,63	0,48 (4,13%)	18.983	18.997	-14 (-0,07%)
1413-8123	84	87	-3 (-3,45%)	2.442	2.315	127 (5,49%)	29,07	26,61	2,46 (9,24%)	71.386	71.512	-126 (-0,18%)
1135-5727	96	96	0	835	805	30 (3,73%)	8,70	8,39	0,31 (3,69%)	24.944	23.697	1.247 (5,26%)
1020-4989	184	186	-2 (-1,08%)	1.674	1.769	-95 (-5,37%)	9,10	9,51	-0,41 (-4,31%)	48.705	49.858	-1.153 (-2,31%)
0864-3466	56	25	31 (124,00%)	610	334	276 (82,63%)	10,89	13,36	-2,47 (-18,49%)	11.879	7.147	4.732 (66,21%)
0213-9111	69	70	-1 (-1,43%)	1.006	851	155 (18,21%)	14,58	12,16	2,42 (19,90%)	24.235	24.534	-299 (-1,22%)
0124-0064	53	47	6 (12,77%)	628	575	53 (9,22%)	11,85	12,23	-0,38 (-3,11%)	15.391	13.949	1.442 (10,34%)
0102-311x	197	218	-21 (-9,63%)	4.001	3.955	46 (1,16%)	20,31	18,14	2,17 (11,96%)	114.408	115.388	-980 (-0,85%)
0042-9686	159	159	0	1.702	1.428	274 (19,19%)	10,70	8,98	1,72 (19,15%)	46.674	47.047	-373 (-0,79%)
0036-3634	80	98	-18 (-18,37%)	1.131	1.014	117 (11,54%)	14,14	10,35	3,79 (36,62%)	34.792	35.064	-272 (-0,78%)
0034-8910	253	261	-8 (-3,07%)	3.428	3.486	-58 (-1,66%)	13,55	13,36	0,19 (1,42%)	66.333	67.157	-824 (-1,23%)

Fonte: desenvolvida pelo autor³⁰

³⁰ Disponível em: < http://cmca.srv.br/prototipo/metabuscador_mostraresumo.php>. Acesso em: 05 abr. 2013.

Três situações identificadas merecem destaque para explicar algumas das diferenças apresentadas entre os números do SciELO e os obtidos na interpretação dos arquivos XML: (I) a exclusão de informações anuais dos dados fonte do SciELO; (II) a estrutura incompleta de *tags* no arquivo XML, que resultou na não incorporação das citações; e (III) problemas na configuração das *tags* dos arquivos XML que inviabilizam o acesso aos mesmos.

As duas primeiras situações foram encontradas para o ISSN 0124-0064, da Revista de Salud Pública. É possível observar que foram interpretados 628 arquivos XML, com 15.391 citações no total:

FIGURA 36 – Estrutura das tags XML dos arquivos interpretados: ISSN 0124-0064

ESTRUTURA DOS ARTIGOS NA BC (tags XML) - 0124-0064	ARTIGOS BC	CITAÇÕES BC
article; front; /front; back; ref-list; ref-id; /ref; /ref-list; /back; /article;	628	15.391
TOTAL	628	15.391

Fonte: desenvolvida pelo autor³¹

Entretanto, o resumo anual do SciELO apresentou 575 artigos e 13.949 citações, conforme a FIG. 37 a seguir. Uma explicação parcial dessa diferença é a ausência dos dados fonte para os anos de 2001 e 2002 (nesse caso, o programa apresenta todos os números do SciELO zerados). É importante ressaltar que, uma vez que existem arquivos XML interpretados para esses dois anos, deduz-se que os dados estavam disponíveis em algum momento (pois o método de importação dos arquivos XML depende dos dados fonte) e foram excluídos.

³¹ Disponível em: < http://cmca.srv.br/prototipo/metabuscaador_mostraisn.php?issn=0124-0064>. Acesso em: 05 abr. 2013.

FIGURA 37 – Resumo da importação de dados do SciELO: dados fonte X arquivos XML: ISSN 0124-0064

0124-0064	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	53	47	6 (12,77%)	628	575	53 (9,22%)	11,85	12,23	-0,38 (-3,11%)	15.391	13.949	1.442 (10,34%)
2012	3	4	-1 (-25,00%)	50	50	0	16,67	12,50	4,17 (33,36%)	1.228	1.228	0
2011	6	6	0	86	87	-1 (-1,15%)	14,33	14,50	-0,17 (-1,17%)	2.082	2.099	-17 (-0,81%)
2010	7	8	-1 (-12,50%)	101	100	1 (1,00%)	14,43	12,50	1,93 (15,44%)	2.605	2.625	-20 (-0,76%)
2009	6	6	0	89	89	0	14,83	14,83	0,00	2.208	2.208	0
2008	6	6	0	88	88	0	14,67	14,67	0,00	2.190	2.190	0
2007	4	4	0	54	54	0	13,50	13,50	0,00	1.168	1.172	-4 (-0,34%)
2006	4	5	-1 (-20,00%)	52	52	0	13,00	10,40	2,60 (25,00%)	1.246	1.003	243 (24,23%)
2005	3	3	0	28	28	0	9,33	9,33	0,00	646	646	0
2004	4	4	0	22	22	0	5,50	5,50	0,00	660	660	0
2003	3	1	2 (200,00%)	18	5	13 (260,00%)	6,00	5,00	1,00 (20,00%)	482	118	364 (308,47%)
2002	4	0	4	23	0	23	5,75	0,00	5,75	410	0	410
2001	3	0	3	17	0	17	5,67	0,00	5,67	466	0	466

Fonte: desenvolvida pelo autor³²

³² Disponível em: < http://cmca.srv.br/prototipo/metabuscaador_mostraiissn.php?issn=0124-0064>. Acesso em: 05 abr. 2013.

S0036-36341998000100004³⁶, S0036-36341999000200008³⁷ e S0036-36342000000200007³⁸. Para todos esses arquivos, a seguinte mensagem foi apresentada:

FIGURA 39 – Erro de acesso a arquivos XML: ISSN 0036-3634



Fonte: SciELO, 2013³⁹

É importante ressaltar que nova consulta realizada no dia 05/04/13 não mais identificou os dados fonte para o período de 1997 a 2000 no SciELO. Dessa forma, nova análise mostrou que as diferenças para o referido periódico foram reduzidas:

³⁶ Disponível em: <<http://www.scielo.org.mx/scieloOrg/php/articleXML.php?pid=S0036-36341998000100004&lang=pt>>. Acesso em: 01 abr. 2013.

³⁷ Disponível em: <<http://www.scielo.org.mx/scieloOrg/php/articleXML.php?pid=S0036-36341999000200008&lang=pt>>. Acesso em: 01 abr. 2013.

³⁸ Disponível em: <<http://www.scielo.org.mx/scieloOrg/php/articleXML.php?pid=S0036-36342000000200007&lang=pt>>. Acesso em: 01 abr. 2013.

³⁹ Disponível em: <<http://www.scielo.org.mx/scieloOrg/php/articleXML.php?pid=S0036-36342000000200007&lang=pt>>. Acesso em: 01 abr. 2013.

FIGURA 40 – Resumo da importação de dados do SciELO: dados fonte X arquivos XML: ISSN 0036-3634

0036-3634	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	80	98	-18 (-18,37%)	1.131	1.014	117 (11,54%)	14,14	10,35	3,79 (36,63%)	34.792	35.064	-272 (-0,78%)
2012	6	6	0	78	61	17 (27,87%)	13,00	10,17	2,83 (27,83%)	2.217	2.217	0
2011	7	10	-3 (-30,00%)	133	101	32 (31,68%)	19,00	10,10	8,90 (88,12%)	3.769	3.769	0
2010	7	8	-1 (-12,50%)	105	93	12 (12,90%)	15,00	11,62	3,38 (29,09%)	2.691	2.963	-272 (-9,18%)
2009	7	10	-3 (-30,00%)	145	119	26 (21,85%)	20,71	11,90	8,81 (74,03%)	5.088	5.088	0
2008	7	10	-3 (-30,00%)	135	108	27 (25,00%)	19,29	10,80	8,49 (78,61%)	3.966	3.966	0
2007	7	10	-3 (-30,00%)	101	97	4 (4,12%)	14,43	9,70	4,73 (48,76%)	3.117	3.117	0
2006	7	8	-1 (-12,50%)	85	85	0	12,14	10,62	1,52 (14,31%)	2.832	2.832	0
2005	6	6	0	46	46	0	7,67	7,67	0,00	1.472	1.472	0
2004	6	6	0	45	45	0	7,50	7,50	0,00	1.496	1.496	0
2003	7	11	-4 (-36,36%)	120	120	0	17,14	10,91	6,23 (57,10%)	3.802	3.802	0
2002	7	7	0	78	78	0	11,14	11,14	0,00	2.409	2.409	0
2001	6	6	0	60	61	-1 (-1,64%)	10,00	10,17	-0,17 (-1,67%)	1.933	1.933	0

Fonte: desenvolvida pelo autor⁴⁰

⁴⁰ Disponível em: < http://cmca.srv.br/prototipo/metabuscador_mostraiissn.php?issn=0036-3634>. Acesso em: 05 abr. 2013.

6.2 Descrição da amostra: a Coleção Saúde Pública em números

Para exemplificar o uso da base de citações, a seguir são apresentados dados quantitativos gerados a partir da importação automática das informações do SciELO para a Coleção Saúde Pública, conforme a metodologia proposta.

Em termos de quantidade de registros, a Coleção Saúde Pública apresentou os seguintes valores para cada tabela do banco de dados, em relação aos periódicos citantes: 14 periódicos, 14 editores, 1.335 fascículos (tabela *edicao*), 23.780 artigos (18.693 apresentaram citações associadas a eles; 5.087 não) e 491.739 citações. Foram encontrados 37.124 resumos (tabela *resumoartigo*) – 10.200 em português, 17.506 em inglês, 8.010 em espanhol e 1.408 em francês – e 44.696 títulos (tabela *tituloartigo*), dos quais 11.804 em português, 20.912 em inglês, 10.563 em espanhol, 1.416 em francês e 1 em latim.

Do total de 149.874 palavras-chave (tabela *palavrachave*) apresentadas em todos os artigos, 35.586 correspondem a termos distintos (tabela *palavra*) – sem nenhum tratamento de desambiguação – que foram apresentados em inglês (15.777), em português (9.201), em francês (1.543) e em espanhol (9.065). As 10 palavras que mais ocorreram nesta amostra foram: México (1.642), Epidemiologia e *Epidemiology* (637 cada), *Risk Factors* (615), *Mortality* (488), *Socioeconomic Factors* (475), *Public health* (474), Colombia (462), Brasil (444) e *Brazil* (438).

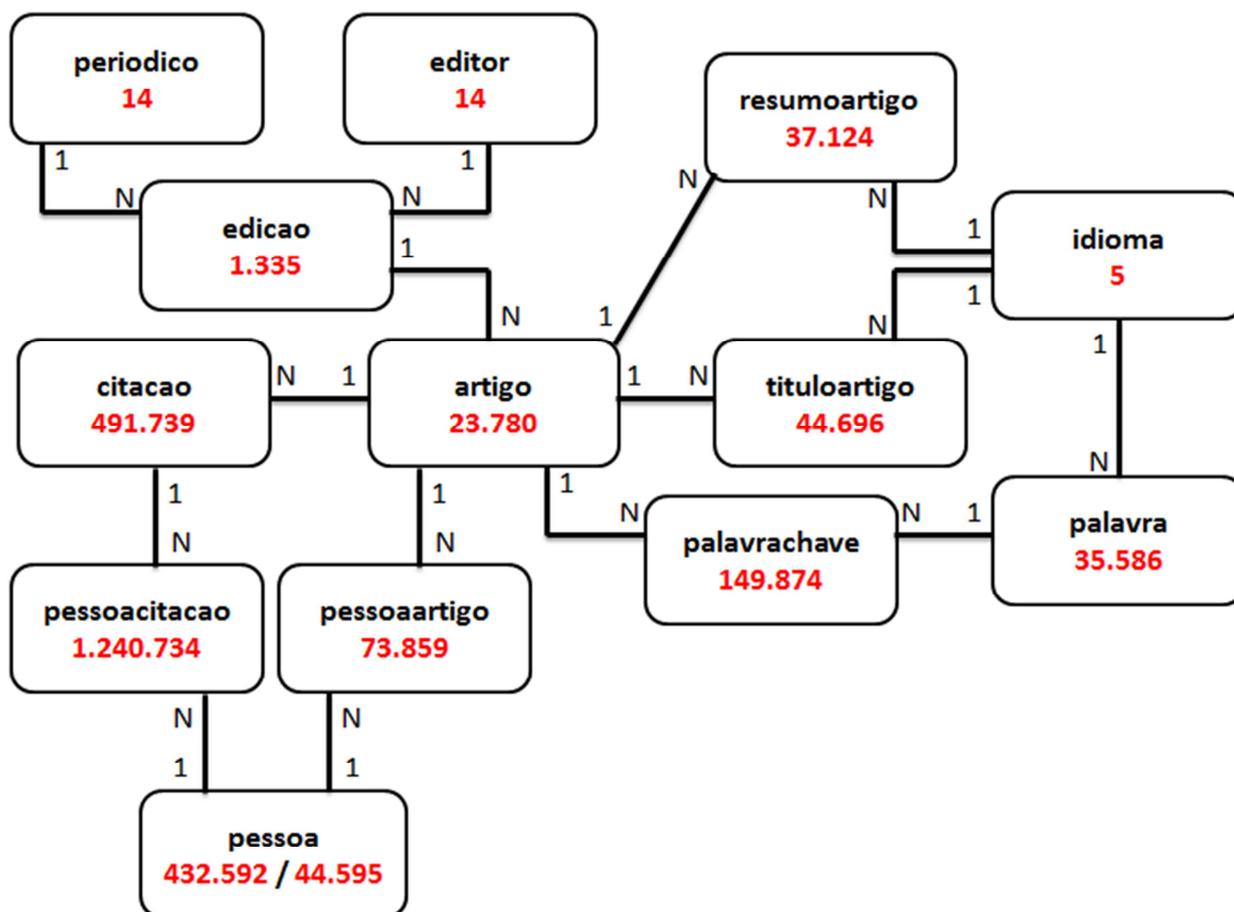
Foram identificados 73.859 autores de artigos (tabela *pessoaartigo*) sendo 44.595 nomes distintos (tabela *pessoa*) – sem desambiguação de nomes. Os 10 autores com mais artigos produzidos foram: Forattini, Oswaldo Paulo (128 artigos), Minayo, Maria Cecília de Souza (100), Victora, Cesar G. (80), Monteiro, Carlos Augusto (79), Leal, Maria do Carmo (77), Laurenti, Ruy (73), Szwarcwald, Celia Landmann (68), Lima-Costa, Maria Fernanda (66), Barros, Marilisa Berti de Azevedo (65), Tomasi, Elaine (63).

Em relação aos periódicos citados, dos 1.240.734 autores identificados nas citações (tabela *pessoacitacao*), 432.592 são distintos (tabela *pessoa*) – sem tratamento de desambiguação. Os 10 nomes mais citados foram: Victora, CG (1.884), Minayo, MCS (1.553), Monteiro, CA (1.278), Barros, FC (997), Szwarcwald, CL (833), Lopez, AD (697), Murray, CJL (667), Lima-Costa, MF (623), Souza, ER (599) e Leal, MC (595).

As fontes mais citadas, desconsiderado o valor “branco” (4.411 ocorrências) e sem nenhum tipo de recorte, foram: Cad Saúde Pública (10.281), *Lancet* (7.222), Rev Saúde Pública (6.596), JAMA (3.482), BMJ (3.455), *Soc Sci Med* (2.472), *Bull World Health Organ*

(2.277), *Am J Public Health* (2.210), *N Engl J Med* (2.123) e *Salud Pública Mex* (2.019). A FIG. 41 apresenta os totais de registros em cada tabela do banco de dados:

FIGURA 41 – Quantidade de registros incorporados no banco de dados: Coleção Saúde Pública



Fonte: desenvolvida pelo autor⁴¹

As próximas FIGURAS deste tópico descrevem a amostra nos formatos dos relatórios apresentados por Garfield (1972, p.527-30): frequências de citações (FIG. 42), estatísticas dos periódicos citados (FIG. 43) e estatísticas dos periódicos citantes (FIG. 44). Para todos os relatórios, foram considerados os 10 periódicos mais citados listados anteriormente e detalhados os 10 últimos anos, com o total dos anos anteriores acumulados na última coluna.

O primeiro acumula o número de vezes que uma referência foi citada, e distribui essas citações por ano em que foram citadas. O segundo, similar ao primeiro, detalha para cada fonte citada os periódicos citantes. Foi usada somente uma amostra parcial do relatório, com o periódico mais citado. A terceira e última lista produzida é similar à segunda, entretanto organiza os dados por periódico citante, detalhando os periódicos citados.

⁴¹ Consulta ao banco de dados realizada em 07 abr. 2013.

FIGURA 42 – Frequências de citações: Coleção Saúde Pública

FONTE CITADA	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Cad Saúde Pública	10.281	280	1.486	1.729	1.281	1.207	1.326	874	788	548	404	358
Lancet	7.222	111	628	795	607	560	630	452	423	388	366	2.262
Rev Saúde Pública	6.596	151	969	932	637	602	748	557	432	422	409	737
JAMA	3.482	58	236	354	257	300	345	268	242	205	216	1.001
BMJ	3.455	36	239	332	304	365	371	270	296	245	299	698
Soc Sci Med	2.472	43	233	266	214	215	307	225	275	198	151	345
Bull World Health Organ	2.277	65	243	250	275	247	278	218	227	86	71	317
Am J Public Health	2.210	31	206	254	193	249	208	200	177	135	135	422
N Engl J Med	2.123	24	191	221	148	248	235	154	146	119	123	514
Salud Pública Mex	2.019	15	154	194	208	253	246	204	205	116	103	321

Fonte: desenvolvida pelo autor⁴²

⁴² Consulta ao banco de dados realizada em 10 abr. 2013.

FIGURA 43 – Estatística dos periódicos citados: Coleção Saúde Pública – visão parcial

FONTE CITADA	CITANTE	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Cad Saude Publica		10.281	280	1.486	1.729	1.281	1.207	1.326	874	788	548	404	358
Cad. Saúde Pública		5.118	140	533	707	521	700	767	576	466	389	319	0
Ciênc. saúde coletiva		2.664	125	614	724	463	240	254	130	107	7	-	0
Rev. Saúde Pública		979	-	118	90	86	85	91	39	145	96	45	184
Rev. bras. epidemiol.		972	-	157	131	124	117	160	87	46	41	30	79
Rev Panam Salud Publica		341	13	51	54	49	24	36	24	7	4	6	73
Rev. salud pública		65	-	4	2	19	23	7	2	6	2	-	0
Salud pública Méx		62	-	1	11	6	14	8	8	7	-	3	4
Gac Sanit		24	-	3	8	1	1	1	3	2	3	-	2
Rev. Esp. Salud Publica		23	-	1	-	-	-	-	1	1	4	1	15
Bull World Health Organ		17	2	-	-	4	3	2	4	1	1	-	0
Rev Peru Med Exp Salud Publica		12	-	3	2	7	-	-	-	-	-	-	0
Rev Cubana Salud Pública		3	-	-	-	1	-	-	-	-	1	-	1
MEDICC rev.		1	-	1	-	-	-	-	-	-	-	-	0

Fonte: desenvolvida pelo autor⁴³

⁴³ Consulta ao banco de dados realizada em 10 abr. 2013.

FIGURA 44 – Estatística dos periódicos citantes: Coleção Saúde Pública – visão parcial

CITANTE	FONTE CITADA	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Cad. Saúde Pública		114.408	1.867	6.844	8.752	6.872	9.945	10.974	10.405	7.845	5.923	6.180	38.801
	Cad Saúde Pública	5.118	140	533	707	521	700	767	576	466	389	319	0
	Rev Saúde Pública	2.988	90	321	427	290	389	467	322	244	217	221	0
	Ciênc Saúde Coletiva	1.192	37	193	194	138	151	183	114	59	77	46	0
	Lancet	1.157	31	75	115	71	94	123	127	57	49	89	326
	Cadernos de Saúde Pública	941	-	-	-	-	-	-	-	-	-	-	941
	Revista de Saúde Pública	803	-	-	-	-	-	-	-	-	-	-	803
	Soc Sci Med	785	24	45	93	69	82	95	102	105	102	68	0
	JAMA	739	5	43	63	47	65	87	85	68	52	53	171
	BMJ	720	7	34	55	49	82	93	78	72	35	45	170
	Diário Oficial da União	708	1	46	62	63	106	138	141	61	21	26	43
	Outros	99.257	1.532	5.554	7.036	5.624	8.276	9.021	8.860	6.713	4.981	5.313	36.347

Fonte: desenvolvida pelo autor⁴⁴

⁴⁴ Consulta ao banco de dados realizada em 10 abr. 2013.

De forma similar aos relatórios apresentados anteriormente, as FIGURAS a seguir apresentam os autores mais citados (FIG. 45), as palavras-chave mais usadas (FIG. 46) e os autores que mais produziram artigos no periódico (FIG. 47). Foram considerados os 10 autores mais citados, classificados em ordem alfabética.

FIGURA 45 – Autores mais citados: Coleção Saúde Pública

AUTOR CITADO	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
Victoria, CG	1.988	57	163	215	215	146	379	136	157	126	110	284
Minayo, MCS	1.634	87	228	180	193	211	167	104	160	104	60	140
Monteiro, CA	1.329	44	152	180	154	150	111	112	56	58	83	229
Barros, FC	1.065	26	84	105	98	54	232	68	89	64	75	170
Szwarcwald, CL	866	41	99	132	81	74	94	103	94	56	49	43
Lopez, AD	710	8	63	62	84	76	65	47	48	47	37	173
Murray, CJL	681	2	32	47	51	72	34	70	51	48	53	221
Souza, ER	660	65	86	64	64	100	57	27	99	28	13	57
Lima-Costa, MF	644	17	107	166	62	71	65	51	31	40	28	6
Leal, MC	607	21	80	71	64	53	64	63	53	49	53	36

Fonte: desenvolvida pelo autor⁴⁵

⁴⁵ Consulta ao banco de dados realizada em 10 abr. 2013.

FIGURA 46 – Palavras-chave mais utilizadas: Coleção Saúde Pública

PALAVRA CITADA	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
México (Espanhol)	818	2	67	59	77	94	79	88	70	42	42	198
Mexico (Inglês)	817	2	65	61	76	96	76	89	68	42	45	197
Epidemiology (Inglês)	654	14	43	54	58	43	40	26	35	22	23	296
Risk factors (Inglês)	629	9	55	42	34	48	62	45	38	52	51	193
Mortality (Inglês)	493	13	35	40	30	41	29	20	34	35	29	187
socioeconomic factors (Inglês)	492	1	34	33	34	49	38	27	23	32	36	185
Public health (Inglês)	478	5	44	52	48	39	37	21	32	19	22	159
Brazil (Inglês)	470	6	47	58	47	63	62	57	32	14	15	69
Epidemiologia (Português)	447	9	27	31	28	19	23	15	14	11	15	255
Fatores de risco (Português)	429	7	37	34	27	37	50	35	29	25	26	122

Fonte: desenvolvida pelo autor⁴⁶

⁴⁶ Consulta ao banco de dados realizada em 10 abr. 2013.

FIGURA 47 – Autores que mais produziram: Coleção Saúde Pública (visão parcial)

CITANTE	AUTOR PRODUTOR	TOTAL	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	ANTERIORES
	Cad. Saúde Pública (0102-311X)	14.822	810	1.023	1.122	1.050	1.257	1.434	1.288	963	762	712	4.401
	Lima-Costa, Maria Fernanda	49	<u>2</u>	<u>2</u>	<u>17</u>	<u>2</u>	<u>4</u>	<u>1</u>	<u>5</u>	<u>3</u>	<u>4</u>	<u>4</u>	5
	Szwarcwald, Celia Landmann	48	<u>1</u>	<u>1</u>	<u>5</u>	<u>1</u>	<u>2</u>	<u>2</u>	<u>1</u>	-	<u>11</u>	<u>4</u>	20
	Victora, Cesar G.	45	-	-	-	<u>6</u>	<u>1</u>	<u>13</u>	-	-	-	<u>2</u>	23
	Leal, Maria do Carmo	44	<u>1</u>	<u>3</u>	<u>2</u>	-	-	<u>2</u>	<u>1</u>	<u>5</u>	<u>2</u>	<u>12</u>	16
	Barros, Fernando C.	37	<u>1</u>	-	<u>1</u>	<u>4</u>	-	<u>11</u>	<u>1</u>	-	<u>1</u>	-	18
	Barros, Marilisa Berti de Azevedo	37	<u>5</u>	<u>8</u>	<u>9</u>	<u>2</u>	<u>3</u>	<u>2</u>	<u>2</u>	<u>3</u>	-	-	3
	Carvalho, Marília Sá	36	<u>2</u>	-	<u>4</u>	<u>1</u>	<u>1</u>	<u>2</u>	<u>4</u>	<u>4</u>	<u>4</u>	<u>5</u>	9
	Olinto, Maria Teresa Anselmo	34	<u>1</u>	<u>2</u>	<u>3</u>	<u>6</u>	<u>2</u>	<u>4</u>	<u>5</u>	<u>5</u>	<u>2</u>	-	4
	Hallal, Pedro C.	33	-	-	<u>2</u>	<u>20</u>	<u>5</u>	<u>2</u>	<u>2</u>	-	<u>2</u>	-	0
	Santos, Iná S.	33	<u>3</u>	-	<u>2</u>	<u>2</u>	<u>3</u>	<u>14</u>	<u>3</u>	-	<u>3</u>	<u>1</u>	2
	Outros	14.426	794	1.007	1.077	1.006	1.236	1.381	1.264	943	733	684	4.301

Fonte: desenvolvida pelo autor⁴⁷

⁴⁷ Consulta ao banco de dados realizada em 10 abr. 2013.

7. ANÁLISE DOS RESULTADOS

Retomando o potencial papel do SciELO como um importante candidato ao trono, apresentado no Capítulo 2, a seguir discute-se o processo de importação dos arquivos XML do SciELO para a geração da base de citações, e como os problemas identificados nos Capítulos anteriores em relação à qualidade da informação no SciELO poderiam ser tratados.

7.1 Análise da obtenção dos dados do SciELO

O processo de obtenção automática dos dados do SciELO foi considerado satisfatório, uma vez que os totais de artigos em cada coleção e os dados fonte anuais foram conferidos manualmente para todas as 10 coleções e não houve diferença em relação aos números apurados a partir dos arquivos XML. Esses dados tem sido atualizados diariamente desde fevereiro do corrente ano, e todas as alterações observadas no SciELO foram refletidas nos dados armazenados.

Entretanto, a diferença apontada no Capítulo 4 – em relação aos dados fonte diferentes para um mesmo ISSN em duas Coleções distintas – pede uma análise que será mais detalhada. Não foi desenvolvida nenhuma análise estatística refinada, pois o grande problema identificado foi a simples existência dessa diferença: parte-se do pressuposto que a menor diferença é problemática, por dois motivos principais: o ISSN é o mesmo, e as Coleções avaliadas são certificadas pelo SciELO⁴⁸, isto é, seguem sua metodologia corretamente.

A análise dos dados fonte apresentados no Capítulo 4 é apresentada a seguir de duas formas: na primeira, foram identificadas as diferenças relativas entre o total – de fascículos, artigos, citações concedidas e recebidas – da coleção de origem e o total na Coleção Saúde Pública; em seguida, são apresentadas as comparações dos dados fonte com os dados extraídos dos arquivos XML para cada um dos 9 periódicos analisados.

As TABELAS 12 a 16 a seguir apresentam os dados da seguinte forma: nas colunas, são apresentados os ISSN que apresentaram alguma diferença para o item avaliado (total anual de fascículos, artigos, citações concedidas ou citações recebidas) e as linhas apresentam cada ano em que houve diferença para qualquer dos ISSN. As diferenças relativas são apresentadas. Por exemplo, conforme os dados da TAB. 12, para o ISSN

⁴⁸ Disponível em <http://www.scielo.org/php/level.php?lang=pt&component=56&item=5>. Acesso em: 19 jun. 2013.

0864-3466, no ano de 1997, a Coleção de Cuba (coleção de origem) apresentou 2 fascículos, e a Coleção Saúde Pública (CSP) não apresentou nenhum. Assim, houve uma diferença de 100% dos artigos a mais na Coleção de origem. No ano de 2011, esse mesmo ISSN apresentou 4 fascículos para a CSP, e 5 para a Coleção de Cuba, gerando uma diferença de 20%:

TABELA 12 – Resumo das diferenças entre o total de fascículos: dados fontes do SciELO

	0036-3634	0124-0064	0213-9111	0864-3466	1135-5727	1413-8123
1.997	-100,0%			100,0%		
1.998	-100,0%			100,0%		
1.999	-100,0%			100,0%		
2.000	-100,0%			100,0%		
2.001		100,0%		100,0%		
2.002		100,0%		100,0%		
2.003		66,7%		100,0%		
2.004			12,5%	100,0%		
2.005				100,0%		
2.006				100,0%		
2.011				20,0%		
2.012	-16,7%	20,0%				
2.013	-100,0%		100,0%	100,0%	100,0%	16,7%

Fonte: desenvolvida pelo autor

Observa-se que para todos os ISSN listados – os demais não apresentaram diferença em relação ao total de fascículos nas duas coleções avaliadas – quando a diferença ocorreu, ela foi positiva, significando que a CSP sempre apresentou número de fascículos menor que a coleção de origem.

A exceção foi o ISSN 0036-3634, que apresentou informações para a CSP nos anos de 1997 a 2000 e 2013 – enquanto a coleção de origem, especificamente México, não apresentou os dados fonte para esses anos, e ainda apresentou um número menor de fascículos (6) em 2012 em relação à CSP (que apresentou 7).

Em relação ao total de artigos, conforme a TAB. 13, também ocorreu a situação de a CSP apresentar dados maiores que a coleção de origem, não somente para o ISSN 0036-3634, mas também para o 0864-3466 (ano de 2007) e para o 0102-311X (anos de 1987 e 2003). Novamente, ressalta-se que qualquer diferença é relevante, pois o ISSN é o mesmo:

TABELA 13 – Resumo das diferenças entre o total de artigos: dados fontes do SciELO

	0036- 3634	0124- 0064	0213- 9111	0864- 3466	1135- 5727	1413- 8123	0102- 311X
1.987							-4,2%
1.995				100,0%			
1.996				100,0%			
1.997	-100,0%			100,0%			
1.998	-100,0%			100,0%			
1.999	-100,0%			100,0%			
2.000	-100,0%			100,0%			
2.001		100,0%		100,0%			
2.002		100,0%		100,0%			
2.003		72,2%	1,4%	100,0%			-0,4%
2.004			8,8%	100,0%			
2.005				100,0%			
2.006				100,0%			
2.007				-37,5%			0,3%
2.008	3,7%			10,2%			
2.012	-14,8%	21,9%					
2.013	-100,0%		100,0%	100,0%	100,0%	17,2%	

Fonte: desenvolvida pelo autor

Para exemplificar outro ponto identificado, comparando-se a TAB. 12 e a TAB. 13, o ISSN 0864-3466 apresentou diferença no total de artigos para os anos de 2007 e 2008, mas não houve diferença em relação ao total de fascículos para esses anos. Como o mesmo número de fascículos, de um mesmo periódico, no mesmo ano – usando a mesma metodologia – pode apresentar total de artigos diferente?

Questionamento similar coloca-se para, por exemplo, o ISSN 1135-5727, que apresentou o mesmo número de fascículos (TAB. 12) e de artigos (TAB. 13) para o período 1997 a 2012, porém o total de citações concedidas foi diferente para os anos de 2003, 2006, 2007 e 2008:

TABELA 14 – Resumo das diferenças entre o total de citações concedidas: dados fontes do SciELO

	0036- 3634	0124- 0064	0213- 9111	0864- 3466	1135- 5727	1413- 8123	0102- 311X	0034- 8910
1.995				100,0%				
1.996				100,0%				
1.997	-100,0%			100,0%				
1.998	-100,0%			100,0%				
1.999	-100,0%			100,0%				
2.000	-100,0%			100,0%				
2.001		100,0%		100,0%				
2.002		100,0%	-0,1%	100,0%				
2.003		75,5%	0,3%	100,0%	1,4%		-1,1%	
2.004			6,3%	100,0%				
2.005				100,0%				
2.006		19,5%	91,6%	100,0%	96,7%			
2.007		-0,3%	-3,4%	-46,2%	-6,5%		0,4%	
2.008	0,1%			-5,7%	-0,9%			
2.009				-15,8%				
2.010		-0,8%						
2.011		0,8%	1,5%	26,7%				
2.012	-14,1%	20,6%						0,1%
2.013	-100,0%		100,0%	100,0%	100,0%	18,4%		

Fonte: desenvolvida pelo autor

As citações recebidas apresentaram a maior diferença relativa entre os ISSN analisados, tanto em termos de valor quanto na frequência:

TABELA 15 – Resumo das diferenças entre o total de citações recebidas: dados fontes do SciELO

	0036- 3634	0124- 0064	0213- 9111	0864- 3466	1135- 5727	1413- 8123	0102- 311X	0034- 8910	1415- 790x
1.980								1,2%	
1.983								1,1%	
1.984								8,2%	
1.985								16,0%	
1.986								21,0%	
1.987							33,3%	10,1%	
1.988								22,2%	
1.989							33,3%	24,1%	
1.990							5,6%	24,2%	
1.991							9,1%	20,5%	
1.992							15,8%	31,1%	
1.993							6,5%	15,4%	
1.994							11,4%	27,5%	
1.995				100,0%			16,3%	22,7%	
1.996				100,0%			22,9%	36,3%	
1.997	-100,0%			100,0%			24,0%	23,8%	
1.998	-100,0%			100,0%		58,3%	23,6%	44,4%	
1.999	-100,0%			100,0%	-2,0%	18,2%	25,5%	47,1%	25,0%
2.000	-100,0%			100,0%	8,3%	47,1%	28,0%	39,3%	22,2%
2.001	28,7%	100,0%		100,0%	26,3%	31,1%	39,4%	51,6%	7,7%
2.002	29,1%	100,0%	8,4%	100,0%	9,6%	20,3%	32,7%	58,0%	30,8%
2.003	18,8%	50,0%	14,1%	100,0%	34,9%	27,1%	36,6%	48,0%	26,1%
2.004	52,9%	31,3%	49,5%	100,0%	50,0%	31,4%	39,7%	49,5%	36,8%
2.005	45,5%	48,0%	22,7%	100,0%	49,5%	35,8%	41,3%	53,2%	34,1%
2.006	46,7%	60,4%	92,5%	100,0%	83,5%	52,5%	47,9%	59,5%	39,2%
2.007	41,3%	68,9%	24,7%	62,3%	42,6%	58,2%	49,3%	55,8%	40,1%
2.008	51,9%	58,8%	29,0%	66,9%	44,4%	53,7%	49,6%	55,0%	46,9%
2.009	46,6%	71,0%	32,0%	72,4%	44,8%	71,7%	55,7%	59,9%	55,2%
2.010	53,8%	66,8%	35,3%	75,3%	54,3%	75,6%	59,4%	63,7%	57,5%
2.011	57,3%	80,3%	42,2%	77,7%	68,7%	73,4%	53,4%	58,8%	54,1%
2.012	61,4%	87,1%	42,7%	70,5%	65,6%	75,3%	59,9%	63,5%	56,7%
2.013	-100,0%		100,0%	100,0%	100,0%	80,3%	54,2%	59,8%	40,8%

Fonte: desenvolvida pelo autor

As diferenças observadas para todas essas informações sugerem a necessidade de revisão da metodologia SciELO implantada para as coleções analisadas. Para tentar identificar se o problema poderia estar na CSP ou em cada coleção de origem, buscou-se a comparação dos valores com mais uma fonte: os arquivos XML do SciELO.

A segunda análise comparou os dados estatísticos diferentes para cada ISSN, apresentados no Capítulo 4, com os dados obtidos a partir dos arquivos XML do SciELO. Neste caso, as citações recebidas não são passíveis de análise, uma vez que não estão disponíveis nos arquivos XML. A TAB. 16 apresenta os totais de cada Coleção e os valores obtidos a partir da importação dos arquivos XML:

TABELA 16 – Periódicos encontrados em mais de uma Coleção do SciELO e dados XML

ISSN	COLEÇÃO	FASCÍCULOS	ARTIGOS	CITAÇÕES CONCEDIDAS
0036-3634	México	98	1.014	35.064
	Saúde Pública	127	1.279	43.387
	XML	80	1.131	34.792
0124-0064	Colômbia	57	644	15.759
	Saúde Pública	47	575	13.949
	XML	54	643	15.726
0213-9111	Espanha	71	866	24.963
	Saúde Pública	68	826	22.552
	XML	69	1.006	24.235
0864-3466	Cuba	60	666	12.279
	Saúde Pública	25	334	7.147
	XML	57	612	11.886
1135-5727	Espanha	97	811	25.235
	Saúde Pública	96	805	23.697
	XML	96	835	24.944
1413-8123	Brasil	91	2.435	75.377
	Saúde Pública	90	2.405	74.338
	XML	87	2.546	74.338
0102-311X	Brasil	220	3.990	116.679
	Saúde Pública	220	3.991	116.703
	XML	199	4.053	116.024
0034-8910	Brasil	263	3.511	67.724
	Saúde Pública	263	3.511	67.720
	XML	253	3.473	67.157
1415-790X	Brasil	57	672	19.600
	Saúde Pública	57	672	19.600
	XML	56	670	19.107

Fonte: desenvolvida pelo autor

Como se observa na TAB. 16, os 9 periódicos apresentam algum tipo de divergência em relação ao total dos dados obtidos dos arquivos XML – dados obtidos exclusivamente do SciELO. As consultas aos dados comparativos entre os arquivos XML e dados do SciELO estão disponíveis na FIG. 48:

FIGURA 48 – Dados comparativos da Coleção Saúde Pública

ISSN	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA
TOTAL (14)	1.343	1.419	-76 (-5,36%)	18.989	18.435	554 (3,01%)	14,14	12,99	1,15 (8,85%)	499.221	508.175	-8.954 (-1,76%)
1851-8265	26	26	0	200	165	35 (21,21%)	7,69	6,35	1,34 (21,10%)	5.820	5.833	-13 (-0,22%)
1726-4634	13	13	0	349	302	47 (15,56%)	26,85	23,23	3,62 (15,58%)	7.104	7.104	0
1555-7960	7	7	0	65	52	13 (25,00%)	9,29	7,43	1,86 (25,03%)	1.781	1.781	0
1415-790x	56	57	-1 (-1,75%)	670	671	-1 (-0,15%)	11,96	11,77	0,19 (1,61%)	19.107	19.567	-460 (-2,35%)
1413-8123	87	90	-3 (-3,33%)	2.546	2.405	141 (5,86%)	29,26	26,72	2,54 (9,51%)	74.338	74.338	0
1135-5727	96	96	0	835	805	30 (3,73%)	8,70	8,39	0,31 (3,69%)	24.944	24.955	-11 (-0,04%)
1020-4989	186	186	0	1.693	1.769	-76 (-4,30%)	9,10	9,51	-0,41 (-4,31%)	49.274	49.858	-584 (-1,17%)
0864-3466	57	60	-3 (-5,00%)	612	666	-54 (-8,11%)	10,74	11,10	-0,36 (-3,24%)	11.886	12.279	-393 (-3,20%)
0213-9111	69	70	-1 (-1,43%)	1.006	851	155 (18,21%)	14,58	12,16	2,42 (19,90%)	24.235	24.534	-299 (-1,22%)
0124-0064	54	47	7 (14,89%)	643	575	68 (11,83%)	11,91	12,23	-0,32 (-2,62%)	15.726	13.949	1.777 (12,74%)
0102-311x	199	219	-20 (-9,13%)	4.053	3.972	81 (2,04%)	20,37	18,14	2,23 (12,29%)	116.024	116.027	-3 (-0,00%)
0042-9686	160	160	0	1.713	1.437	276 (19,21%)	10,71	8,98	1,73 (19,27%)	47.033	47.406	-373 (-0,79%)
0036-3634	80	127	-47 (-37,01%)	1.131	1.279	-148 (-11,57%)	14,14	10,07	4,07 (40,42%)	34.792	43.387	-8.595 (-19,81%)
0034-8910	253	261	-8 (-3,07%)	3.473	3.486	-13 (-0,37%)	13,73	13,36	0,37 (2,77%)	67.157	67.157	0

Fonte: consulta ao banco de dados⁴⁹

⁴⁹ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

Com base nos dados apresentados na FIG. 48, e analisando-se as FIGURAS 73 a 86 do Anexo B, observa-se que, mesmo considerando os ISSN que estão vinculados exclusivamente à Coleção Saúde Pública (ou seja, para os 14 ISSN avaliados), ao comparar os números estatísticos do SciELO com os dados extraídos dos arquivos XML:

- a) Mesmo quando existe coerência no total de fascículos (1851-8265, 1726-4634, 1555-7960, 1135-5727, 1020-4989 e 0042-9686) ou citações (1726-4634, 1555-7960, 1413-8123 e 0034-8910), o total de artigos não foi idêntico para nenhum ISSN;
- b) Analisando cada ano de cada ISSN, apenas 29 em 224 (12,95%) estavam idênticos em número de fascículos, artigos e citações concedidas);
- c) O ISSN 0034-8910 apresentou maior percentual de anos com todas as informações idênticas: 50% (23 de 46). Os ISSN 0124-0064, 0864-3466 e 0102-311 apresentaram dados idênticos para, respectivamente: 2 em 12 (16,67%), 3 em 19 (15,79%), e 1 em 29 (3,45%) dos registros de dados fonte anuais.
- d) Todos os outros ISSN/ano apresentaram alguma divergência, principalmente em relação ao total de artigos.

Essas divergências podem ser explicadas, ao menos parcialmente, a partir dos problemas identificados: (I) a diferença entre os dados fonte de um mesmo ISSN em coleções diferentes; (II) a exclusão de informações anuais dos dados fonte do SciELO; (III) a estrutura incompleta de tags no arquivo XML, que resultou na não incorporação das citações; e (IV) problemas na configuração das tags dos arquivos XML que inviabilizaram o acesso aos mesmos

7.2 Análise da criação da base de citações

A importação dos arquivos foi considerada satisfatória, uma vez que durante a realização de testes para o desenvolvimento do protótipo foram obtidos mais de 260 mil arquivos XML, todos eles interpretados e inseridos na base de citações.

As quatro situações identificadas nos capítulos anteriores⁵⁰, caso corrigidas, podem reduzir significativamente a diferença de números entre os dados fonte do SciELO e a importação dos arquivos XML – melhorando a qualidade das informações disponibilizadas.

⁵⁰ (I) a exclusão de informações anuais dos dados fonte do SciELO; (II) a estrutura incompleta de tags no arquivo XML, que resultou na não incorporação das citações; (III) problemas na

Conforme as FIGURAS 49 e 50 a seguir, observa-se que para todos os periódicos (a coleção de Saúde Pública e a PCI), a BC apresentou 3.361 citações a mais que o SciELO, representando 0,69% de um total de 495.966 na consulta realizada em 05/04/2013 (FIG. 50). Uma semana antes, entretanto, havia uma diferença a menor na BC em relação ao total de citações do SciELO: 1,09% ou 5.490 citações em relação a um total de 503.779 (FIG. 49).

FIGURA 49 – Resumo da importação de dados do SciELO: dados fonte X arquivos XML em 31/03/2013

ISSN	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA
TOTAL	1.353	1.398	-45 (-3,22%)	18.934	18.214	720 (3,95%)	13,99	13,03	0,96 (7,37%)	498.289	503.779	-5.490 (-1,09%)
1851-8265	26	26	0	200	165	35 (21,21%)	7,69	6,35	1,34 (21,10%)	5.820	5.833	-13 (-0,22%)
1726-4634	12	12	0	311	278	33 (11,87%)	25,92	23,17	2,75 (11,87%)	6.509	6.509	0
1555-7960	6	6	0	55	46	9 (19,57%)	9,17	7,67	1,50 (19,56%)	1.554	1.560	-6 (-0,38%)
1415-790x	55	56	-1 (-1,79%)	666	651	15 (2,30%)	12,11	11,63	0,48 (4,13%)	18.983	18.997	-14 (-0,07%)
1413-9936	24	24	0	292	290	2 (0,69%)	12,17	12,08	0,09 (0,75%)	7.714	7.714	0
1413-8123	84	87	-3 (-3,45%)	2.414	2.315	99 (4,28%)	28,74	26,61	2,13 (8,00%)	70.701	71.512	-811 (-1,13%)
1135-5727	96	96	0	835	805	30 (3,73%)	8,70	8,39	0,31 (3,69%)	24.944	23.697	1.247 (5,26%)
1020-4989	184	186	-2 (-1,08%)	1.674	1.769	-95 (-5,37%)	9,10	9,51	-0,41 (-4,31%)	48.705	49.858	-1.153 (-2,31%)
0864-3466	56	25	31 (124,00%)	610	334	276 (82,63%)	10,89	13,36	-2,47 (-18,49%)	11.879	7.147	4.732 (66,21%)
0213-9111	69	70	-1 (-1,43%)	1.006	851	155 (18,21%)	14,58	12,16	2,42 (19,90%)	24.235	24.534	-299 (-1,22%)
0124-0064	53	47	6 (12,77%)	628	575	53 (9,22%)	11,85	12,23	-0,38 (-3,11%)	15.391	13.949	1.442 (10,34%)
0102-311x	197	218	-21 (-9,63%)	4.000	3.955	45 (1,14%)	20,30	18,14	2,16 (11,91%)	114.393	115.388	-995 (-0,86%)
0042-9686	159	159	0	1.701	1.428	273 (19,12%)	10,70	8,98	1,72 (19,15%)	46.671	47.047	-376 (-0,80%)
0036-3634	80	126	-46 (-36,51%)	1.131	1.271	-140 (-11,01%)	14,14	10,09	4,05 (40,14%)	34.792	42.976	-8.184 (-19,04%)
0034-8910	252	260	-8 (-3,08%)	3.411	3.481	-70 (-2,01%)	13,54	13,39	0,15 (1,12%)	65.998	67.058	-1.060 (-1,58%)

Fonte: desenvolvida pelo autor⁵¹

⁵¹ Disponível em: <http://cmca.srv.br/prototipo/metabuscar_mostraresumo.php>. Acesso em: 31 mar. 2013.

FIGURA 50 – Resumo da importação de dados do SciELO: dados fonte X arquivos XML em 05/04/2013

ISSN	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA
TOTAL (15)	1.354	1.371	-17 (-1,24%)	18.981	17.962	1.019 (5,67%)	14,02	13,10	0,92 (7,02%)	499.327	495.966	3.361 (0,68%)
1851-8265	26	26	0	200	165	35 (21,21%)	7,69	6,35	1,34 (21,10%)	5.820	5.833	-13 (-0,22%)
1726-4634	12	12	0	311	278	33 (11,87%)	25,92	23,17	2,75 (11,87%)	6.509	6.509	0
1555-7960	6	6	0	55	46	9 (19,57%)	9,17	7,67	1,50 (19,56%)	1.554	1.560	-6 (-0,38%)
1415-790x	55	56	-1 (-1,79%)	666	651	15 (2,30%)	12,11	11,63	0,48 (4,13%)	18.983	18.997	-14 (-0,07%)
1413-9936	24	24	0	292	290	2 (0,69%)	12,17	12,08	0,09 (0,75%)	7.714	7.714	0
1413-8123	84	87	-3 (-3,45%)	2.442	2.315	127 (5,49%)	29,07	26,61	2,46 (9,24%)	71.386	71.512	-126 (-0,18%)
1135-5727	96	96	0	835	805	30 (3,73%)	8,70	8,39	0,31 (3,69%)	24.944	23.697	1.247 (5,26%)
1020-4989	184	186	-2 (-1,08%)	1.674	1.769	-95 (-5,37%)	9,10	9,51	-0,41 (-4,31%)	48.705	49.858	-1.153 (-2,31%)
0864-3466	56	25	31 (124,00%)	610	334	276 (82,63%)	10,89	13,36	-2,47 (-18,49%)	11.879	7.147	4.732 (66,21%)
0213-9111	69	70	-1 (-1,43%)	1.006	851	155 (18,21%)	14,58	12,16	2,42 (19,90%)	24.235	24.534	-299 (-1,22%)
0124-0064	53	47	6 (12,77%)	628	575	53 (9,22%)	11,85	12,23	-0,38 (-3,11%)	15.391	13.949	1.442 (10,34%)
0102-311x	197	218	-21 (-9,63%)	4.001	3.955	46 (1,16%)	20,31	18,14	2,17 (11,96%)	114.408	115.388	-980 (-0,85%)
0042-9686	159	159	0	1.702	1.428	274 (19,19%)	10,70	8,98	1,72 (19,15%)	46.674	47.047	-373 (-0,79%)
0036-3634	80	98	-18 (-18,37%)	1.131	1.014	117 (11,54%)	14,14	10,35	3,79 (36,62%)	34.792	35.064	-272 (-0,78%)
0034-8910	253	261	-8 (-3,07%)	3.428	3.486	-58 (-1,66%)	13,55	13,36	0,19 (1,42%)	66.333	67.157	-824 (-1,23%)

Fonte: desenvolvida pelo autor⁵²

52

Disponível

em:

http://cmca.srv.br/prototipo/metabuscaador_mostraresumo.php?x=0&ss=640324979&id=5e8eaacaa6791d624b67f40de34f1456>. Acesso em: 05 abr. 2013.

É importante ressaltar que os números dos dados fonte do SciELO foram reduzidos nesse intervalo - indicando que houve exclusão de dados fonte no SciELO para os ISSN estudados. No caso do ISSN 0036-3634, essa medida parece ter corrigido o problema de erro no acesso aos arquivos XML para o período entre 1997 e 2000 – provavelmente os arquivos realmente não estavam disponíveis. Observa-se que, no período, o número de fascículos do SciELO caiu de 126 para 98; de artigos, de 1.271 para 1.014, e de citações, de 42.976 para 35.064. Com essa redução, a diferença entre os dados da BC e do SciELO caiu de 8.184 (19,04%) para 272 (0,78%).

Sugere-se uma análise mais aprofundada da forma como o SciELO gera os dados fonte, pois mesmo com o baixo percentual de diferença entre o total de citações, as discrepâncias entre o total de fascículos e artigos encontradas são representativas e podem gerar divergências nos cálculos de indicadores a partir dos dados fonte do SciELO – e esses indicadores, por sua vez, podem ser usados para subsidiar políticas de avaliação de periódicos.

Um exemplo pode ser observado na FIG. 51, para o ano de 2003, que apresenta o mesmo número de artigos e citações no SciELO e na BC, porém a BC registra 4 fascículos a menos – ou seja, os 120 artigos e citações da BC foram obtidos a partir de arquivos XML distribuídos em apenas 7 fascículos, e não 11 como apontam os dados fonte do SciELO.

FIGURA 51 – Resumo da importação de dados do SciELO para o ISSN 0036-3634

0036-3634	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	80	98	-18 (-18,37%)	1.131	1.014	117 (11,54%)	14,14	10,35	3,79 (36,63%)	34.792	35.064	-272 (-0,78%)
2012	6	6	0	78	61	17 (27,87%)	13,00	10,17	2,83 (27,83%)	2.217	2.217	0
2011	7	10	-3 (-30,00%)	133	101	32 (31,68%)	19,00	10,10	8,90 (88,12%)	3.769	3.769	0
2010	7	8	-1 (-12,50%)	105	93	12 (12,90%)	15,00	11,62	3,38 (29,09%)	2.691	2.963	-272 (-9,18%)
2009	7	10	-3 (-30,00%)	145	119	26 (21,85%)	20,71	11,90	8,81 (74,03%)	5.088	5.088	0
2008	7	10	-3 (-30,00%)	135	108	27 (25,00%)	19,29	10,80	8,49 (78,61%)	3.966	3.966	0
2007	7	10	-3 (-30,00%)	101	97	4 (4,12%)	14,43	9,70	4,73 (48,76%)	3.117	3.117	0
2006	7	8	-1 (-12,50%)	85	85	0	12,14	10,62	1,52 (14,31%)	2.832	2.832	0
2005	6	6	0	46	46	0	7,67	7,67	0,00	1.472	1.472	0
2004	6	6	0	45	45	0	7,50	7,50	0,00	1.496	1.496	0
2003	7	11	-4 (-36,36%)	120	120	0	17,14	10,91	6,23 (57,10%)	3.802	3.802	0
2002	7	7	0	78	78	0	11,14	11,14	0,00	2.409	2.409	0
2001	6	6	0	60	61	-1 (-1,64%)	10,00	10,17	-0,17 (-1,67%)	1.933	1.933	0

Fonte: desenvolvida pelo autor⁵³

⁵³ Disponível em: < http://www.cmca.srv.br/prototipo/metabuscaador_mostraisn.php?issn=0036-3634&ss=1288361369&id=5e8eaacaa6791d624b67f40de34f1456>. Acesso em: 05 abr. 2013.

Ainda de acordo com as FIGURAS 49 e 50, os ISSN 0042-9686, 0102-311X e 1413-8123 apresentaram alterações positivas apenas na BC em relação ao número de artigos (e, conseqüentemente, na média) e de citações, indicando que houve a incorporação de novos arquivos XML pelo sistema. Os ISSN citados apresentaram, respectivamente para cada ISSN, 1, 1 e 28 arquivos e 3, 15 e 685 citações a mais em relação à consulta realizada no dia 31/03/2013.

No mesmo período o ISSN 0034-8910 apresentou crescimento de 1 fascículo, 5 artigos e 99 citações, de acordo com os dados do SciELO. Entretanto, a BC registrou um aumento de 1 fascículo, 17 artigos e 335 citações. A FIG. 52 a seguir comprova os dados da BC. Todos os outros ISSN apresentados nas FIGURAS 49 e 50, tanto em relação aos dados do SciELO quanto da BC, permaneceram inalterados no período.

FIGURA 52 – Artigos incorporados na BC: ISSN 0034-8910 no período entre 31/03 e 05/04/2013

Artigos: 17

#	Autoria	Ano	Periódico	Título 	Volume	Número	Páginas	Referências			
1	GOLDBAUM, Moisés FRANÇA JÚNIOR, Ivan SOUZA, Luis Eugenio Portela Fernandes de	2012	Rev. Saúde Pública		46		1-2	1			
2	GADELHA, Carlos Augusto Grabois VARGAS, Marco COSTA, Lais Silveira MALDONADO, José	2012	Rev. Saúde Pública	A dinâmica inovativa do subsistema de base mecânica, eletrônica e de materiais	46		29-36	22			
3	GADELHA, Carlos Augusto Grabois VARGAS, Marco COSTA, Lais Silveira MALDONADO, José BORGES, Tais Raiher BURD, Paula	2012	Rev. Saúde Pública	A dinâmica inovativa para a reestruturação dos serviços de saúde	46		76-82	26			
4	GADELHA, Carlos Augusto Grabois COSTA, Lais Silveira MALDONADO, José	2012	Rev. Saúde Pública	A perspectiva territorial da inovação em saúde: a necessidade de um novo enfoque	46		59-67	38			
5	SILVA, Hudson P PETRAMALE, Clarice A ELIAS, Flavia T S	2012	Rev. Saúde Pública	Avanços e desafios da política nacional de gestão de tecnologias em saúde	46		83-90	13			
6	GADELHA, Carlos Augusto Grabois COSTA, Lais Silveira BRITTO, Jorge VARGAS, Marco Antônio	2012	Rev. Saúde Pública	Competências científico- tecnológicas e cooperação universidade-empresa na saúde	46		41-50	10			
7	CHOR, Dóra BARRETO, Sandhi Maria SCHMIDT, Maria Inês DUNCAN, Bruce Bartholow MILL, José Geraldo LOTUFO, Paulo Andrade BENSEÑOR, Isabela M VIGO, Álvaro AQUINO, Estela M L	2012	Rev. Saúde Pública	Doenças crônicas não transmissíveis no Brasil: prioridade para enfrentamento e investigação	46		126-134	33			

Fonte: desenvolvida pelo autor⁵⁴

⁵⁴ Disponível em: < >. Acesso em: 05 abr. 2013.

FIGURA 52 – Artigos incorporados na BC: ISSN 0034-8910 no período entre 31/03 e 05/04/2013 (cont.)

Artigos: 17

#	Autoria	Ano	Periódico	Título 	Volume	Número	Páginas	Referências			
8	RIVERA, Francisco Javier Uribe ARTMANN, Elizabeth MARTINS, Wagner de Jesus	2012	Rev. Saúde Pública	Gestão comunicativa para redes cooperativas de ciência, tecnologia e inovação em saúde	46		51-58	30			
9	COSTA, Karen Sarmento NASCIMENTO JR., José Miguel do	2012	Rev. Saúde Pública	HÓRUS: inovação tecnológica na assistência farmacêutica no sistema único de saúde	46		91-99	24			
10	GADELHA, Carlos Augusto Grabois VARGAS, Marco COSTA, Lais Silveira MALDONADO, José	2012	Rev. Saúde Pública	Inovação na indústria química e biotecnológica em saúde: em busca de uma agenda virtuosa	46		37-40	4			
11	FACCHINI, Luiz Augusto MATOS, Divane Leite LIMA-COSTA, Maria Fernanda MACINKO, James	2012	Rev. Saúde Pública	Mudanças em dez anos das desigualdades sociais em saúde dos idosos brasileiros (1998-2008)	46		100-107	21			
12	CONDE, Wolney Lisboa KUPEK, Emil CAMPOS, Vanessa Caroline ASSIS, Maria Alice Altenburg de COSTA, Filipe Ferreira da LEAL, Danielle Biazzi	2012	Rev. Saúde Pública	Mudanças no consumo alimentar e atividade física de escolares de Florianópolis, SC, 2002 - 2007	46		117-125	25			
13	GADELHA, Carlos Augusto Grabois COSTA, Lais Silveira MALDONADO, José	2012	Rev. Saúde Pública	O complexo econômico-industrial da saúde e a dimensão social e econômica do desenvolvimento	46		21-28	5			
14	BARBOSA, Pedro Ribeiro GADELHA, Carlos Augusto Grabois MOURA, Erly Catarina de GOMES, Romeu	2012	Rev. Saúde Pública	O papel dos hospitais na dinâmica de inovação em saúde	46		68-75	19			
15	COUTO, Márcia Thereza SCHWARZ, Eduardo CARVALHO, Sarah de Araújo SILVA, Simione Fátima Cesar da	2012	Rev. Saúde Pública	Política de saúde do homem	46		108-116	37			
16	GADELHA, Carlos Augusto Grabois COSTA, Lais Silveira	2012	Rev. Saúde Pública	Saúde e desenvolvimento no Brasil: avanços e desafios	46		13-20	24			
17	GADELHA, Carlos Augusto Grabois	2012	Rev. Saúde Pública	Saúde e desenvolvimento: uma nova abordagem para uma nova política	46		5-8	3			

Fonte: desenvolvida pelo autor

7.3 Falta de padronização de nomes de autores e fontes

Os dados detalhados coletados a partir dos arquivos XML oferecem diversas possibilidades de análise bibliométrica, que extrapolam os objetivos aqui propostos.

Entretanto, observou-se que o primeiro passo para essa análise deve ser a revisão criteriosa dos dados levantados buscando a uniformização e normalização dos mesmos, pois foram detectadas diferenças importantes de grafia, por exemplo, do nome das fontes: “Persp. Ci. Inf.”, “Perspect. Ciênc. Inf.”, “Perspectiva em Ciência da Informação”, “Perspectivas em Ciência da Informação” e “Perspectivas em Ciências da Informação”; “*Annual Review of Information Science and Technology*” e “*Annual Review of Information Science and Technology - ARIST*”; “Ci. Inf”, “Ci. Inf.”, “Ciência da Informação” e “Ciências da Informação”; “Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.”, “Encontros Bibli” e “Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação”; “*Information Processing & Management*” e “*Information Processing and Management*”, entre outros.

Os nomes de autores também necessitam de desambiguação, uma vez que a grafia é diferente, a exemplo dos resultados apresentados nas FIGURAS 53 e 54 a seguir. A pesquisa por autor foi realizada usando o termo ‘BOCCATO’. A FIG. 53 apresenta os artigos produzidos pela autora, e a FIG. 54 os artigos em que ela é citada:

FIGURA 53 – Consulta artigos produzidos pelo autor BOCCATO na PCI

# Aatoria	Ano	Periódico	Título 	Volume	Número	
1	BOCCATO, Vera Regina Casari	2010	Perspect. ciênc. inf.		15	3
2	BOCCATO, Vera Regina Casari FUJITA, Mariângela Spotti Lopes	2010	Perspect. ciênc. inf.	O uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo de avaliação sociocognitiva com protocolo verbal	15	3
3	BOCCATO, Vera Regina Casari LIMA, Vania Mara Alves	2009	Perspect. ciênc. inf.	O desempenho terminológico dos descritores em Ciência da Informação do Vocabulário Controlado do SIBi/USP nos processos de indexação manual, automática e semi-automática	14	1
4	BOCCATO, Vera Regina Casari FUJITA, Mariângela Spotti Lopes	2006	Perspect. ciênc. inf.	Estudos de avaliação quantitativa e qualitativa de linguagens documentárias: uma síntese bibliográfica	11	2

Fonte: desenvolvida pelo autor⁵⁵

Percebe-se que houve um padrão para os 4 artigos de autoria de BOCCATO na PCI: BOCCATO, Vera Regina Casari. Entretanto, essa grafia não se repetiu em nenhuma das 14 citações identificadas na PCI para a autora, sendo as outras: BOCCATO, V. R. C.; BOCCATO, V. R. C; BOCCATO, V. R C.; BOCCATO, BOCCATO, V.R.C. e BOCCATO, V. C.:

⁵⁵ Disponível em:
<http://www.cmca.srv.br/prototipo/metabuscaador_busca.php?ss=1373015907&id=e912bb7fea529b66434443659672ad78>. Acesso em: 12 jun. 2013.

FIGURA 54 – Consulta pelo autor BOCCATO nas citações da PCI

#	Autoria	Ano	Título	Fonte	Tipo	Citado por	Ano	Periódico	Artigo 
1	BOCCATO, V. R. C FUJITA, M. S. L.	2011	Estudo comparativo entre vocabulários controlados de catálogos coletivos em bibliotecas universitárias.	Anais...	confpro	FUJITA, Mariângela Spotti Lopes LACRUZ, Maria del Carmen Agustin DÍAZ, Raquel Gómez	2012	Perspect. ciênc. inf.	A situação atual da indexação nas tarefas bibliotecárias
2	BOCCATO, V. R. C.			Avaliação do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo sociocognitivo com protocolo verbal		FUJITA, Mariângela Spotti Lopes LACRUZ, Maria del Carmen Agustin DÍAZ, Raquel Gómez	2012	Perspect. ciênc. inf.	A situação atual da indexação nas tarefas bibliotecárias
3	FUJITA, M. S. L. RUBI, M. P. BOCCATO, V. C.	2009	O contexto sociocognitivo do catalogador em bibliotecas universitárias: perspectivas para uma política de tratamento da informação documentária	Datagramazero	journal	FUJITA, Mariângela Spotti Lopes LACRUZ, Maria del Carmen Agustin DÍAZ, Raquel Gómez	2012	Perspect. ciênc. inf.	A situação atual da indexação nas tarefas bibliotecárias
4	FUJITA, M. S. L. RUBI, M. P. BOCCATO, V. R. C.	abr	O contexto sociocognitivo do catalogador em bibliotecas universitárias: perspectivas para uma política de tratamento da informação documentária	DataGramaZero	journal	DAL' EVEDOVE, Paula Regina FUJITA, Mariângela Spotti Lopes	2012	Perspect. ciênc. inf.	Teoria e prática em catalogação de assunto: a sistematicidade do processo em contexto de bibliotecas universitárias pela perspectiva profissional
5	BOCCATO, V. R. C.	2009		Avaliação do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo sociocognitivo com protocolo verbal		BOCCATO, Vera Regina Casari	2010	Perspect. ciênc. inf.	
6	BOCCATO, V. R. C. FUJITA, M. S. L.	2006	A atividade de indexação nas perspectivas das concepções de assunto: o protocolo verbal como instrumento de avaliação qualitativa-cognitiva	Anais...	confpro	BOCCATO, Vera Regina Casari	2010	Perspect. ciênc. inf.	
7	BOCCATO, V. R. C.	2005		Avaliação de linguagem documentária em Fonoaudiologia na perspectiva do usuário: estudo de observação da recuperação da informação com protocolo verbal		BOCCATO, Vera Regina Casari FUJITA, Mariângela Spotti Lopes	2010	Perspect. ciênc. inf.	O uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo de avaliação sociocognitiva com protocolo verbal

Fonte: desenvolvida pelo autor

FIGURA 54 – Consulta pelo autor BOCCATO nas citações da PCI (cont.)

#	Autoria	Ano	Título	Fonte	Tipo	Citado por	Ano	Periódico	Artigo 
8	BOCCATO, V. R. C.	2009		Avaliação do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo sociocognitivo com protocolo verbal		 BOCCATO, Vera Regina Casari FUJITA, Mariângela Spotti Lopes	2010	Perspect. ciênc. inf.	O uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo de avaliação sociocognitiva com protocolo verbal
9	BOCCATO, V.R.C. FUJITA, M.S.L.	2006	Avaliação da linguagem documentária DeCS na área de fonoaudiologia na perspectiva do usuário: estudo de observação da recuperação da informação com protocolo verbal	Encontros Bibli	journal	 VOGEL, Michely Jabala Mamede	2009	Perspect. ciênc. inf.	A influência da Jean-Claude Gardin e a linha francesa na evolução do conceito de linguagem documentária
10	BOCCATO, V. R. C.	2005		Avaliação de linguagem documentária em Fonoaudiologia na perspectiva do usuário: estudo de observação da recuperação da informação com protocolo verbal		 LIMA, Vania Mara Alves BOCCATO, Vera Regina Casari	2009	Perspect. ciênc. inf.	O desempenho terminológico dos descritores em Ciência da Informação do Vocabulário Controlado do SIBi/USP nos processos de indexação manual, automática e semi-automática
11	BOCCATO, V. R. C.	2005		Avaliação de linguagem documentária em Fonoaudiologia na perspectiva do usuário: estudo de observação da recuperação da informação com protocolo verbal		 LIMA, Vania Mara Alves BOCCATO, Vera Regina Casari	2009	Perspect. ciênc. inf.	O desempenho terminológico dos descritores em Ciência da Informação do Vocabulário Controlado do SIBi/USP nos processos de indexação manual, automática e semi-automática
12	LIMA, V. M. A. BOCCATO, V. R. C.	mai/	Atualização da lista de assuntos USP: compatibilização de linguagens documentárias	Ciência da Informação	journal	 LIMA, Vania Mara Alves BOCCATO, Vera Regina Casari	2009	Perspect. ciênc. inf.	O desempenho terminológico dos descritores em Ciência da Informação do Vocabulário Controlado do SIBi/USP nos processos de indexação manual, automática e semi-automática

Fonte: desenvolvida pelo autor

FIGURA 54 – Consulta pelo autor BOCCATO nas citações da PCI (cont.)

#	Autoria	Ano	Título	Fonte	Tipo	Citado por	Ano	Periódico	Artigo 
13	BOCCATO, V. R. C.	2005		Avaliação de linguagem documentária em Fonoaudiologia na perspectiva do usuário: estudo de observação da recuperação da informação com protocolo verbal		 BOCCATO, Vera Regina Casari FUJITA, Mariângela Spotti Lopes	2006	Perspect. ciênc. inf.	Estudos de avaliação quantitativa e qualitativa de linguagens documentárias: uma síntese bibliográfica
14	CARVALHO, T. BOCCATO, V. R. C. RAMOS, L. M. S. V. C.	1998	Estudo do usuário da Sub-Rede Nacional de informação na Área de Ciências da Saúde Oral, sob o enfoque do "Sense-Making"	Anales...	confpro	 BOCCATO, Vera Regina Casari FUJITA, Mariângela Spotti Lopes	2006	Perspect. ciênc. inf.	Estudos de avaliação quantitativa e qualitativa de linguagens documentárias: uma síntese bibliográfica

Fonte: desenvolvida pelo autor⁵⁶

⁵⁶ Disponível em: <http://www.cmca.srv.br/prototipo/metabuscaador_busca.php?ss=1238359810&id=70cef30ff360861295f5b2e750aa2a5b>. Acesso em: 12 jun. 2013.

A FIG. 54 também permite a identificação de conteúdos alfanuméricos na coluna ANO (citações 4 e 12). A correção desse conteúdo no processo de geração dos arquivos XML é importante para que as consultas agrupadas por ano, e mesmo qualquer outra consulta referenciando o ano da citação, não sejam comprometidas.

8. CONCLUSÕES E RECOMENDAÇÕES

A metodologia proposta mostrou-se capaz de gerar corretamente a base de citações a partir dos arquivos XML. Em alguns casos foi possível a identificação de 4 tipos de problemas: estrutura de *tags* do XML incompleta, impedindo a importação das citações; arquivos XML não acessíveis, com erro de estrutura XML; dados fonte de periódicos que estavam disponíveis no início do projeto mas foram excluídos do SciELO (nesses casos, os arquivos XML incorporados geravam diferenças). Mais detalhes sobre a forma como o SciELO calcula os dados sobre número de fascículos, artigos e citações precisam ser investigados para a análise das diferenças encontradas.

Considerando o SciELO uma fonte secundária de informações, percebe-se a fragilidade desta pesquisa em relação ao conteúdo das informações geradas para a composição da base de citações. Uma forma sugerida para contornar essa fragilidade trata da revisão de políticas editoriais para garantir a qualidade dos dados primários em cada periódico – o que aumentaria a qualidade das informações do SciELO, e, conseqüentemente, da base de citações criada.

Outra sugestão inclui a adoção de gerenciadores de citações para a edição dos artigos submetidos a um periódico, garantindo a maior padronização das informações, e também a geração automática da base de citações, o que dispensaria a utilização de uma fonte secundária.

Além disso, sugere-se a inclusão nos arquivos XML da seção à qual cada artigo pertence, em determinada edição, para que futuras análises possam filtrar e interpretar esses artigos – a exemplo de artigos de revisão ou inéditos – e realizar interpretações mais específicas.

Para a consecução deste trabalho, a exemplo de outros já citados, foi demonstrada a importância da interdisciplinaridade entre a Ciência da Informação e a Ciência da Computação, o que permitiu inclusive a geração de um produto informacional que poderá ser usado para o desenvolvimento e suporte de inúmeras pesquisas.

A seguir são apresentados os próximos passos identificados para a melhoria do protótipo, algumas sugestões para a realização de futuras pesquisas e a relevância do trabalho realizado do ponto de vista prático, acadêmico e de gestão.

Cabe ressaltar que, por se tratar de uma metodologia única para todas as coleções do SciELO, outros países que possuem (ou que venham a possuir) coleções cadastradas no SciELO poderão se beneficiar da mesma forma, desde que mantidos os padrões identificados para os endereços das coleções, dados fonte e arquivos XML.

8.1 Próximos passos

Inicialmente, apesar dos problemas identificados quanto à correção da informação, a continuação desta pesquisa prevê a criação de bases de citações específicas que poderão ser instaladas na página de acesso a cada um dos periódicos tratados, que poderão oferecer diversas consultas bibliométricas a exemplo dos relatórios de autores mais citados, estatísticas de periódicos citantes e de periódicos citados.

Também é prevista a criação de uma base de citações centralizada, com todos os periódicos nacionais listados na Coleção Brasil, oferecendo consultas mais abrangentes em relação à produção científica nacional. Outra iniciativa é a preparação de um servidor específico para a criação de uma base de citações com todas as coleções do SciELO citadas neste trabalho.

Para estas bases (de cada periódico e centralizada) está em estudo a criação de uma facilidade que permita, a partir de uma lista de artigos, indicar quantas citações existem em cada um deles e quais os outros artigos que citam as mesmas referências – sendo esta talvez a função mais útil, do ponto de vista de utilização pelo pesquisador, de uma base de citações.

Para a gestão do conteúdo das bases citadas, pretende-se criar interfaces que permitam a desambiguação manual dos nomes de autores e fontes – esse processo poderia ser vinculado à prática dos cursos de graduação da ECI-UFMG, ou discutido com a própria equipe do SciELO.

A criação dessas bases poderá fornecer subsídios para novas pesquisas, entre elas as que estão destacadas no próximo tópico.

8.2 Futuras pesquisas

Sugere-se, inicialmente, a realização de duas pesquisas: a primeira relacionada à desambiguação automática de nomes de fontes e autores; a outra, à proposição de um processo automático de verificação da disponibilidade das referências citadas no SciELO no PORTAL CAPES – auxiliando o processo de gestão de seu acervo.

Ainda que pesem questões políticas e culturais para a implantação dessas sugestões, pretende-se fomentar a reflexão sobre a viabilidade e os benefícios advindos dessas propostas.

8.2.1 Desambiguação automática de nomes de autores e fontes

Seguindo a mesma ideia de automação de processos para facilitar a análise e composição da base de citações, discute-se a seguir uma alternativa para a desambiguação automática dos nomes de autores e das fontes citadas.

É importante ressaltar que a geração das bases de citação – do SciELO, ou de qualquer outra coleção digital – por si só não garante a correção de inúmeros problemas em relação ao seu conteúdo, o que pode inviabilizar a sua utilização para a gestão do conhecimento científico. Por esse motivo a integração de alguns Sistemas de Informação em Ciência e Tecnologia (SICT) é discutida a seguir para viabilizar a desambiguação automática de nomes em duas situações: nome dos autores, e nome da fonte de cada citação.

A premissa fundamental para a discussão deste cenário tem como base a modelagem essencial (MATTOS, 2012b): cadastros únicos. Quando a informação está centralizada, todos os sistemas podem acessá-la e a posterior compatibilização de informações fica assegurada no processo de recuperação. É importante ressaltar que a centralização aqui definida não se vincula exclusivamente ao domínio restrito ou exclusividade de posse e atualização; a perspectiva proposta é de centralização do armazenamento das informações, e não necessariamente da alteração de seu conteúdo. Um exemplo é a plataforma LATTES: a informação está centralizada, porém cada pesquisador é responsável pela atualização de seus dados.

Em relação à desambiguação de nomes de autores, um ponto estratégico é a centralização de um cadastro – no caso dos pesquisadores brasileiros, a plataforma LATTES. A partir deste cadastro unificado, os periódicos OJS/SEER, a Biblioteca Digital de Teses e Dissertações (BDTD) e outros sistemas podem buscar as informações por CPF, eliminando a necessidade de desambiguação posterior e facilitando a recuperação de informações para cada autor nos sistemas citados. Mesmo os problemas de homônimos estariam resolvidos com o uso do CPF. O mesmo vale para sistemas de gerenciamento de eventos.

Não foram realizados experimentos nesse sentido, mas uma sugestão é a informação do CPF do autor ao cadastrar-se um artigo em periódico OJS – a plataforma faria uma consulta automática à plataforma LATTES, preenchendo os dados do autor a partir das informações retornadas. Dessa forma, toda vez que o autor fosse cadastrado em um periódico OJS, seus dados estariam coerentes.

Para tratar das situações em que os dados do autor são eventualmente alterados, idealmente a plataforma OJS/SEER seria alterada para armazenar o campo CPF do autor, que ficaria armazenado posteriormente no XML gerado pelo SciELO. O tratamento do CPF na geração da base de citações eliminaria a necessidade de desambiguação do nome de autores, e também resolveria o problema de homônimos. Também a própria informação sobre o sexo do autor seria incorporada à análise, o que mostrou-se difícil na análise da literatura (SOTUDEH e KHOSHIAN, 2013).

Em termos internacionais, de forma análoga a plataforma OJS poderia integrar-se ao *ResearcherID* ou ao Google Acadêmico (centralizadores do cadastro de autores), por exemplo – mas a decisão política para essa padronização dependeria do interesse de diversos países. No caso do Brasil, a FAPESP adotou esses cadastros como informação imprescindível à obtenção de financiamento.

Exemplo adicional refere-se à desambiguação do nome dos periódicos: todos eles possuem uma referência internacional: o ISSN. Caso o ISSN fosse informado nas citações, haveria a padronização de nomes de periódicos, facilitando a realização de estudos bibliométricos e o cruzamento de informações entre os periódicos.

Outro ponto essencial é a centralização da tabela de áreas e subáreas de conhecimento, bem como da de instituições de ensino. LATTES, BDTD e OJS/SEER poderiam consultar estas tabelas, cuja alteração ficaria a cargo da agência responsável. A princípio, esses cadastros existem – falta a definição política para a sua adoção em vista dos benefícios para a gestão da informação em ciência e tecnologia. Um exemplo é o WebQualis, que concentra periódicos nacionais (ISSN) avaliados em áreas e subáreas de conhecimento – e poderia também concentrar a instituição responsável de cada periódico. Nesse caso, os outros SICT poderiam consultar esse conjunto de dados unificado, garantindo maior padronização – condição que aumenta a possibilidade de integração e recuperação de informações. O PORTAL OJS proposto por Mattos (no prelo) apresenta a estrutura integrada de áreas e subáreas de conhecimento, periódicos (ISSN) e estrato QUALIS de avaliação para periódicos nacionais da Ciência da Informação.

8.2.2 Avaliação automática de coleções digitais: uma proposta para o PORTAL CAPES

Entre as bibliotecas digitais relacionadas à produção científica nacional destaca-se o PORTAL CAPES. Sua importância deve-se, por um lado, à disponibilidade de um amplo acervo digital a toda a comunidade científica brasileira, e por outro, à disseminação de parte dessa produção em alguns dos periódicos digitais cadastrados.

Conforme o Capítulo 2, observa-se que a gestão do acervo do PORTAL CAPES poderia se beneficiar caso fosse definido um processo automático de verificação da disponibilidade das referências citadas na BC desenvolvida para o SciELO.

A partir da BC criada para a Coleção Saúde Pública, por exemplo, seria possível avaliar automaticamente se as fontes citadas – e as referências usadas – estão disponíveis no acervo do PORTAL CAPES. O complemento também seria possível: quais as fontes citadas nos artigos da Coleção Saúde Pública que não estão disponíveis no PORTAL CAPES?

Sugere-se a realização de novas pesquisas para análise da viabilidade de implantação dessa integração entre o SciELO e o PORTAL CAPES.

8.3 Considerações finais: a relevância do trabalho realizado

A partir da disponibilização das bases de citações para os periódicos do SciELO, este trabalho tornar-se-á relevante academicamente ao permitir o desenvolvimento de uma série de estudos cientométricos, que poderão ser realizados de forma mais rápida e abrangente, eliminando a necessidade de coleta manual dos dados.

Do ponto de vista prático, este trabalho representa um passo inicial importante para a ampliação dos recursos de pesquisa à produção científica nacional, além de ser fonte de dados para políticas científicas e avaliação de pesquisadores, entre outros, fortalecendo o desenvolvimento científico e oferecendo maior visibilidade aos periódicos e autores nacionais.

Em relação à gestão da produção científica, a automação do processo de geração da base de citações oferece recursos que podem contribuir para a melhoria do acervo e visibilidade das revistas científicas. A partir da informação das fontes utilizadas em cada periódico, registrada na base de citações, também será viável a identificação de eventuais problemas de qualidade dos periódicos – a exemplo da recente suspensão de

alguns periódicos nacionais da classificação QUALIS em função de autoplágio e empilhamento de citações.

Apesar de fornecer indicadores estatísticos sobre os periódicos cadastrados em sua base, o SciELO apresenta algumas diferenças que precisam ser investigadas, e ainda carece de uma forma mais automática para que os pesquisadores se utilizem dessas informações.

Dessa forma, talvez a contribuição mais significativa deste trabalho seja a criação de uma metodologia que permitirá a interpretação de todos os arquivos XML disponíveis nas Coleções do SciELO, oferecendo assim a possibilidade para a criação de uma base de citações para a América Latina, Caribe e outras coleções do SciELO. Essa base poderá prover informações para subsidiar políticas e inúmeras pesquisas, destacando-se pela cobertura de diversas áreas de conhecimento.

REFERÊNCIAS

- ADAM, D. Citation analysis: The counting house. *Nature*, v. 415, p. 726-729, 2002. Disponível em: <<http://www.nature.com/nature/journal/v415/n6873/full/415726a.html>>. Acesso em: 20 jun. 2013.
- ALMEIDA, E. C. E; GUIMARÃES, J. A.; ALVES, I. T. G. Dez anos do Portal de Periódicos da Capes: histórico, evolução e utilização. *RBPG*, Brasília, v. 7, n. 13, p. 218 - 246, novembro de 2010. Disponível em: <http://www2.capes.gov.br/rbpg/images/stories/downloads/RBPG/Vol.7_13/1_Artigo.pdf>. Acesso em: 01 set 2012.
- ALVARADO, R. U. A Bibliometria no Brasil. *Ci. Inf.*, Brasília, v. 13, n. 2, p. 91-105, 1984. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1444>>. Acesso em: 30 jun. 2013.
- ALVARADO, R. U. A lei de Lotka e a produtividade dos autores. 2007. 240f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.
- BOSHOF, N. Neo-colonialism and research collaboration in Central Africa. *Scientometrics*, v. 81, n. 2, p. 413–434, 2009. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-008-2211-8.pdf>>. Acesso em 11 ago. 2013.
- BOHMERWALD, P. Uma proposta metodológica para avaliação de bibliotecas digitais: usabilidade e comportamento de busca por informação na Biblioteca Digital da PUC - Minas. *Ci. Inf.*, Brasília, v. 34, n. 1, p. 95-103, jan/abr. 2005. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652005000100011>. Acesso em: 30 jul. 2011.
- BRASIL. Senado Federal. *Produção científica no Brasil: um salto no número de publicações*. Disponível em: <<http://www.senado.gov.br/NOTICIAS/JORNAL/EMDISCUSSAO/inovacao/investimento-inovacao-tecnologica-finep-pesquisadores-brasil/producao-cientifica-no-brasil-um-salto-no-numero-de-publicacoes.aspx>>. Acesso em: 31 dez. 2012.
- BRODY, S. Impact factor: Imperfect but not yet replaceable. *Scientometrics*, v. 96, p. 255-257, 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-012-0863-x.pdf>>. Acesso em: 05 ago. 2013.
- CAMARGO JR, K. R. O rei está nú, mas segue impávido: os abusos da bibliometria na avaliação da ciência. *Saúde & Transformação Social / Health & Social Change [On-line]*, v. 1,

p. 3-8, 2010. Disponível em: <<http://www.redalyc.org/articulo.oa?id=265319560003>>. Acesso em: 05 jul. 2013.

CATHCART, R.; ROBERTS, A.. Evaluating Google Scholar as a tool for information literacy. *Internet Reference Services Quarterly*, v. 10, n. 3–4, p. 167–176, 2005. Disponível em: <http://www.researchgate.net/publication/233468977_Evaluating_Google_Scholar_as_a_Tool_for_Information_Literacy?ev=pub_cit>. Acesso em: 10 ago. 2013.

CENDÓN, B. V.; GUIMARÃES, M. C. S.; SILVA, C. H.; OLIVEIRA, M.; MATTOS, M. C.; SANTANA, R. A. L.; FERNANDES, W. R. *Construção e validação de metodologia e protótipo para criação de um índice de citações da produção científica brasileira: um estudo de caso na área de saúde coletiva*. Projeto submetido à chamada MCTI/CNPQ/MEC/CAPES nº 18/2012. 2012. (Não publicado)

CENDÓN, B. V.; RIBEIRO, N. A. Análise da literatura acadêmica sobre o Portal Periódico CAPES. *Informação & Sociedade: Estudos*, João Pessoa, v. 18, n. 2, p. 157-178, maio/ago. 2008. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000005049&dd1=c6f7d>>. Acesso em: 20 nov 2011.

CHEN, P. Entity-Relationship Modeling: Historical Events, Future Trends, and Lessons Learned. In: *Software Pioneers: Contributions to Software Engineering*, Broy M. and Denert, E. (eds.), Springer-Verlag, Berlin, Lecturing Notes in Computer Sciences, June 2002, pp. 100-114. Disponível em: <http://www.csc.lsu.edu/~chen/pdf/Chen_Pioneers.pdf>. Acesso em 23 fev. 2011.

CHEN, P. The Entity-Relationship Model-Toward a Unified View of Data. *ACM Transactions on Database Systems*, Vol. 1, No. 1. March 1976, Pages 9-36. Disponível em: <http://www.colonese.it/00-Sw-Engineering_Articoli/Entity-Relationship%20Model,%20Peter%20Chen,%201976.pdf>. Acesso em 23 fev. 2011.

CODD, E. F. A relational model of data for large shared data banks. *Commun. ACM [S.I.]*, v. 13, n. 6, p. 377-387, 1970. Disponível em: <<http://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf>>. Acesso em 23 fev. 2011.

CODD, E. F. *Derivability, redundancy and consistency of relations stored in large data banks*. IBM Research Report, 1969. Disponível em: <<http://www.sigmod.org/publications/sigmod-record/0903/p17.40year.codd.pdf>>. Acesso em 23 fev. 2011.

COLLAZO-REYES, F. Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country. *Scientometrics*, on-line, 15 May 2013. Disponível em: <<http://link.springer.com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-1036-2.pdf>>. Acesso em: 05 ago. 2013.

COLLAZO-REYES, F. et al. Publication and citation patterns of Latin American & Caribbean journals in the SCI and SSCI from 1995 to 2004. *Scientometrics*, v. 75, n. 1, p. 145–161, 2008. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-007-1841-6.pdf>>. Acesso em: 05 ago. 2013.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (Capes). Tabela de áreas de conhecimento. Disponível em: <<http://www.capes.gov.br/avaliacao/tabela-de-areas-de-conhecimento>>. Acesso em: 08 nov. 2012.

CORREA, C. H. W. et al. Portal de Periódicos da CAPES: um misto de solução financeira e inovação. *Revista Brasileira de Inovação*, v. 7, p. 127-145, 2008. Disponível em: <<http://www.ige.unicamp.br/ojs/index.php/rbi/article/view/332/251>>. Acesso em: 24 set 2012.

CRESWELL, J. W.; CLARK, V. L. *Designing and conducting mixed methods research*. 2nd. ed. Los Angeles: Sage, 2011. xxvi, 457p.

CUNHA, A. A. L. *Uso de bibliotecas digitais de periódicos: um estudo comparativo no Portal de Periódicos CAPES entre áreas de conhecimento*. 2009. 207 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

DANILOWICZ, C.; SZARSKI, H. Selection of scientific journals based on the data obtained from an information service system. *Information Processing & Management*, v. 17, n. 1, 1981, p. 13-19. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0306457381900376>>. Acesso em: 12 fev 2011.

DELANGHE, H.; SLOAN, B.; MUL DUR, U. European research policy and bibliometric indicators, 1990–2005. *Scientometrics*, v. 87, p. 389-398, 2011. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/article/10.1007/s11192-010-0308-3>>. Acesso em: 10 ago. 2013.

DIAS, G. A. Avaliação do acesso a periódicos eletrônicos na web pela análise do arquivo de log de acesso. *Ci. Inf.*, Brasília, v. 31, n. 1, p. 7-12, jan./abr. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n1/a02v31n1.pdf>>. Acesso em: 12 fev 2011.

DONLAN, R.; COOKE, R. Running with the devil. Accessing library-licensed full text holdings through Google Scholar. *Internet Reference Services Quarterly*, v. 10, n. 3–4, p. 149–157. 2005. Disponível em: <http://www.tandfonline.com/doi/abs/10.1300/J136v10n03_13#.Ug_9rpKh1sk>. Acesso em: 10 ago. 2013.

FERNANDES, W. R. *Ciência da informação no portal de periódicos capes* : estudo bibliométrico para avaliação da coleção e da interdisciplinaridade da área. 2009. 121 f. Monografia (Especialização em Gestão Estratégica da Informação)-Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

FERNANDES, W. R. *PORTAL PERIÓDICOS CAPES*: estudo dos não-usuários docentes das IFES brasileiras. 2012. 263 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

FERNANDES, W. R.; CENDÓN, B. V. Portal de Periódicos da Capes: proposta de um modelo de estudo bibliométrico para avaliação da coleção. *RBPG*, Brasília, v. 7, n. 13, nov. 2010. Disponível em: <http://www2.capes.gov.br/rbpg/images/stories/downloads/RBPG/Vol.7_13/5_Artigo.pdf>. Acesso em: 25 ago. 2011.

GALVEZ, M. Publicaciones biomedicas: realidad de Chile y Latinoamerica. *Rev. chil. radiol.* [online], v.12, n. 3, p.113-117, 2006. Disponível em: <http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0717-93082006000300003&lng=en&nrm=iso>. Acesso em: 12 ago. 2013.

GARFIELD, E. Citation Indexing for Studying Science. *Nature*, v. 227, p. 669-71, 1970. Disponível em: <<http://garfield.library.upenn.edu/essays/V1p133y1962-73.pdf>>. Acesso em: 10 jan. 2013.

GARFIELD, E. Citation analysis as a tool in journal evaluation. *Science*, v. 178, p. 471-479, 1972. Disponível em: <<http://www.garfield.library.upenn.edu/essays/V1p527y1962-73.pdf>>. Acesso em: 10 jan. 2013.

GARFIELD, E. Is citation analysis a legitimate evaluation tool? *Scientometrics*, v. 1, p. 359-375, 1979. Disponível em: <[http://www.garfield.library.upenn.edu/papers/scientometricsv1\(4\)p359y1979.pdf](http://www.garfield.library.upenn.edu/papers/scientometricsv1(4)p359y1979.pdf)>. Acesso em: 10 jan. 2013.

GARFIELD, E. Citation data: their use as quantitative indicators for science and technology evaluation and policy-making. *Science & Public Policy*, v. 19, n. 5, p.321-327, 1992. Disponível em: <[http://www.garfield.library.upenn.edu/papers/sciandpubpolv19\(5\)p321y1992.html](http://www.garfield.library.upenn.edu/papers/sciandpubpolv19(5)p321y1992.html)>. Acesso em: 10 jan. 2013.

GARFIELD, E. Quantitative analysis of the scientific literature and its implications for science policymaking in Latin America and the Caribbean. *Bull Pan Am Health Organ*, v. 29, p. 87-95, 1995. Disponível em: <[http://www.garfield.library.upenn.edu/papers/paho29\(1\)p87y1995.pdf](http://www.garfield.library.upenn.edu/papers/paho29(1)p87y1995.pdf)>. Acesso em: 12 jan. 2013.

GARFIELD, E. Citation Indexes for Retrieval and Research Evaluation. *Consensus Conference on the Theory and Practice of Research Assessment*, Capri, 2006. Disponível em: <<http://www.garfield.library.upenn.edu/papers/ciretreseval-capri.html>>. Acesso em: 10 jan. 2013.

GLÄNZEL, W. *A concise introduction to bibliometrics and its history: the origins of the term bibliometrics*. Policy and Research Centre for R & D indicators. [2003?]. Disponível em: <<http://www.ecoom.be/en/research/bibliometrics>> Acesso em: 20 fev. 2013.

GOGOLIN, I. et al. European social science citation index: a chance for promoting European research? *European Educational Research Journal*, v.2, n.4, p.574-593, 2003. Disponível em: <http://www.worlds.co.uk/pdf/validate.asp?j=eerj&vol=2&issue=4&year=2003&article=6_Roundtable_EERJ_2_4_web>. Acesso em: 20 jan. 2013.

GOLDENBERG, S. et al. Interpretação dos dados estatísticos da SciELO (Scientific Electronic Library Online). *Acta Cir. Bras. [online]*, v. 22, n. 1, [cited 2007-08-10], p. 1-7, 2007. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-86502007000100001&lng=en&nrm=iso>. Acesso em: 12 ago. 2013.

GONZÁLEZ-ALCAIDE, G.; VALDERRAMA-ZURIÁN, J. C.; ALEIXANDRE-BENAVENT, R. The Impact Factor in non-English-speaking countries. *Scientometrics*, v. 92, p. 297-311, 2012. Disponível em: <[http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-012-0692-y.pdf](http://link.springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-012-0692-y.pdf)>. Acesso em: 05 ago. 2013.

GUEDES, R. D. *O Projeto SciELO e os Repositórios Institucionais de textos científicos*. 2012. Dissertação (Mestrado do Programa de Pós-Graduação em Políticas Públicas, Estratégias e Desenvolvimento) - Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012. Disponível em: <<http://objdig.ufrj.br/43/dissert/RodrigoDuarteGuedes.pdf>>. Acesso em: 12 ago. 2013.

GUIMARÃES, M. C. S.; SILVA, C. H.; SANTANA, R. A. L.; BRAGA, G. M.; BOCHNER, R.; GOLDBAUM, M. *Métricas em saúde coletiva: bases quantitativas e qualitativas para a criação de um índice de citação da literatura nacional em Saúde Coletiva*. Relatório de pesquisa para o Projeto CNPq – Processo 403522/2008-0. 2011. (Não publicado)

HARTLEY, J.; CABANAC, G. Do men and women differ in their use of tables and graphs in academic publications?. *Scientometrics*, on-line, 27 July 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-1096-3.pdf>>. Acesso em 10 ago. 2013.

HARZING, A. A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics*, v. ?, p. ?, 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/article/10.1007/s11192-013-0975-y>>. Acesso em: 10 ago. 2013.

JACSÓ, P. Google Scholar revisited. *Online Information Review*, v. 32, n. 1, p. 102–114, 2008. Disponível em: <<http://comminfo.rutgers.edu/~tefko/Courses/e530/Readings/Jaszo%20Google%20scholar%202008.pdf>>. Acesso em: 09 ago. 2013.

JACSÓ, P. Google Scholar: The pros and the cons. *Online Information Review*, v. 29, n. 2, p. 208–214, 2005. Disponível em: <<http://www.emeraldinsight.com/journals.htm?articleid=1464936&show=pdf>>. Acesso em: 10 ago. 2013.

KRAUSKOPF M. et al. A citationist perspective on science in Latin America and The Caribbean, 1981-1993. *Scientometrics*, v. 34, n. 1, p. 3-25, 1995. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2FBF02019169.pdf>>. Acesso em 05 ago. 2013.

KUMAR, S.; JAN, J. M. Research collaboration networks of two OIC nations: comparative study between Turkey and Malaysia in the field of 'Energy Fuels', 2009–2011. *Scientometrics*, on-line, 11 June 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-1059-8.pdf>>. Acesso em 10 ago. 2013.

LAKATOS, E. M.; MARCONI, M. A. *Técnicas de pesquisa*. São Paulo: Atlas, 6ed, 2007. 289 p.

LANCASTER, F. W. *Avaliação de serviços de bibliotecas*. Brasília: Briquet de Lemos Livros, 1996.

LOURENÇO, C. A. L. Modelagem de dados como ferramenta de análise de padrões de metadados em bibliotecas digitais: o padrão de metadados brasileiro para teses e dissertações segundo o modelo entidade-relacionamento. 2005. 161 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/EARM-6ZGNZC/doutorado___c_ntia_de_azevedo_louren_o.pdf?sequence=1>. Acesso em: 25 jan. 2013.

MACIAS-CHAPULA, C. A. Influence of local and regional publications in the production of public health research papers in Latin America. *Scientometrics*, v. 84, p. 703-716, 2010. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-009-0153-4.pdf>>. Acesso em: 05 ago. 2013.

MACIAS-CHAPULA, C. A.; RODEA-CASTRO, I. P.; NARVAEZ-BERTHELEMOT, N. Bibliometric analysis of AIDS literature in Latin America and The Caribbean. *Scientometrics*, v. 41, n. 1-2, p. 41-49, 1998. Disponível em: <<http://link-springer->

com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2F02457965.pdf>. Acesso em: 05 ago. 2013.

MAIA, L. C. G. *Uso de periódicos eletrônicos: um estudo sobre o Portal de Periódicos Capes na Universidade Federal de Minas Gerais*. 2005. 153 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

MAIA, L. C. G.; CENDÓN, B. V. Uso de periódicos eletrônicos: o Portal de Periódicos da Capes na UFMG. *RBPG*, Brasília, v. 9, n. 17, p.425-56, jul. 2012. Disponível em: <http://www2.capes.gov.br/rbpg/images/stories/downloads/RBPG/Vol.9_17/estudo4.pdf>. Acesso em: 23 dez. 2012.

MAIA, R. M. C. S. *Intercomplementação do Sistema Categorical de Ranganathan e do Modelo Entidade-Relacionamento de Chen para a Modelagem Conceitual: uma aplicação no domínio do biomonitoramento do Projeto Manuelzão/UFMG nas águas da bacia do Rio das Velhas*. 2013. [? f]. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

MARCONDES, C. H. et al. (Orgs.). *Bibliotecas digitais: saberes e práticas*. 2. ed. Salvador: EDUFBA ; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2006.

MATTOS, M. C.; CENDÓN, B. V. Da possibilidade de uma *Web of Science* para a América Latina e Caribe: extração automática de uma base de citações do SciELO para o periódico *Perspectivas em Ciência da Informação* e para a Coleção de Saúde Pública. In: 65a Reunião Anual da SBPC – Sociedade Brasileira para o Progresso da Ciência. *Anais...* Recife, 2013. No prelo. Disponível em: <http://www.cmca.srv.br/blogmax/_arquivos/2013-SBPC.pdf>. Acesso em: 31 jul. 2013.

MATTOS, M. C.. O papel da tecnologia na construção do conhecimento. *NAVUS – Revista de Gestão e Tecnologia*. Florianópolis, v. 2, n. 2, p. 33 – 39, jul./dez. 2012a. Disponível em: <<http://navus.sc.senac.br/index.php/navus/article/view/80>>. Acesso em: 02 jan. 2013.

MATTOS, M. C. A redundância implícita na modelagem relacional tradicional: uma breve introdução à modelagem essencial. In: IX CONTECSI – International Conference on Information Systems and Technology Management, 9., 2012b. São Paulo. *Anais...* São Paulo: FEA-USP, 2012. Disponível em: <http://www.cmca.srv.br/blogmax/_arquivos/contecsi2012redundancia.pdf>. Acesso em: 10 dez. 2012.

MATTOS, M. C. Metabuscaador QUALIS OJS: proposta para a criação de portais de periódicos institucionais ou por área de conhecimento. No prelo.

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

MÊGNIGBÊTO, E. International collaboration in scientific publishing: the case of West Africa (2001–2010). *Scientometrics*, on-line, 24 Feb. 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-0963-2.pdf>>. Acesso em 11 ago. 2013.

MENDONÇA, M. A. R. Serviços de referência digital. In: MARCONDES, C. H. et al. (orgs.). *Bibliotecas digitais: saberes e práticas*. Salvador, BA: EDUFBA; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2 ed, 2006. 337 p.

MENEGHINI, R.; PACKER, A. L. Is there science beyond English? Initiatives to increase the quality and visibility of non-English publications might help to break down language barriers in scientific communication. *EMBO Reports*, v. 8, n.2, p. 112-6, 2007. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1796769/>>. Acesso em 05 ago. 2013.

MENEGHINI, R.; MUGNAINI, R.; PACKER, A. L. International versus national oriented Brazilian scientific journals. A scientometric analysis based on SciELO and JCR-ISI databases. *Scientometrics*, v. 69, n. 3, p. 529–538, 2006. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-006-0168-z.pdf>>. Acesso em: 05 ago. 2013.

MENEGHINI, R. Avaliação da produção científica e o Projeto SciELO. *Ci. Inf.*, Brasília, v. 27, n. 2, p. 219-220, 1998. Disponível em: <<http://www.scielo.br/pdf/ci/v27n2/meneghini.pdf>>. Acesso em: 15 jan. 2013.

MEYER, M. et al. Origin and emergence of entrepreneurship as a research field. *Scientometrics*, on-line, 31 May 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-1021-9.pdf>>. Acesso em 10 ago. 2013.

MOTTA, D. F. Validade da análise de citação como indicador de qualidade da produção científica: uma revisão. *Ciência da Informação*, Brasília, v.12, n.1, p.53-59, 1983. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1540/1163>>. Acesso em: 05 jun. 2013.

MUELLER, S. P. M. Metodologia para avaliação de lista básica de periódicos. *Ciência da Informação*, v. 20, n. 2, 1991, p. 111-118. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/viewFile/1227/868>>. Acesso em: 13 mar. 2012.

MUGNAINI, Rogério. *Caminhos para adequação da avaliação da produção científica brasileira: impacto nacional versus internacional*. 2006. Tese (Doutorado em Cultura e Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2006. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-11052007-091052/>>. Acesso em: 25 abr. 2013.

NEGISHI, M.; SUN, Y.; SHIGI, K. Citation Database for Japanese Papers: a new bibliometric tool for Japanese academic society. *Scientometrics*, v. 60, p. 333-351, 2004. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1023%2FB%3ASCIE.0000034378.38698.b2.pdf>>. Acesso em 05 ago. 2013.

NOURMOHAMMADI, H.; HODAEI, F. Perspective of Iranian women's scientific production in high priority fields of science and technology. *Scientometrics*, on-line, 02 Aug. 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-1098-1.pdf>>. Acesso em: 10 ago. 2013.

OLIVEIRA, M. A. F.; ODDONE, N. Estudo bibliométrico sobre o uso do acervo do portal da capes entre os pesquisadores brasileiros da ciência da informação. In: *ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 8., 2007, Salvador. [Anais eletrônicos]... Salvador: UFBA, 2007

PACKER, A. L. et al. SciELO: una metodología para la publicación electrónica. *Revista Española de Salud Pública*, v.75, n.4, 2001. Disponível em: <http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S1135-57272001000400004&lng=en&nrm=iso>. Acesso em: 12 ago. 2013.

PACKER, A. L. Os periódicos brasileiros e a comunicação da pesquisa nacional. *Rev. USP [online]*, n. 89, p. 26-61, 2011. Disponível em:<http://rusp.scielo.br/scielo.php?script=sci_arttext&pid=S0103-99892011000200004&lng=en&nrm=iso>. Acesso em: 12 ago. 2013.

PACKER, A. L. et al. SciELO: uma metodologia para publicação eletrônica. *Ciência da Informação [online]*,v. 27, n. 2, 1998. Disponível em: <www.scielo.br/cgi-bin/fbpe/fbtext?got=last&pid=S0100-19651998000200002&usr=fbpe&lng=en&seq=0100-1965-005&nrm=iso&sss=1&aut=71981947>. Acesso em: 12 ago. 2013.

POBLACIÓN, D. A.; GOLDENBERG, S. Acta Cirúrgica Brasileira. Visibilidade e acessibilidade da produção científica na área da cirurgia experimental. *Acta Cirúrgica Brasileira [online]*, v. 16, n. 3, 2001. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-86502001000300001&lng=en&nrm=iso>. Acesso em: 12 ago. 2013.

PRAT, A. M. Programa Biblioteca Científica Electrónica en Línea, SciELO-Chile: una nueva forma de acceder a la literatura científica nacional. *Biological Research [online]*, v. 33, n. 2, 2000. Disponível em: <http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0716-97602000000200003&lng=es&nrm=iso>. Acesso em: 12 ago. 2013.

PRATHAP, G. Big data and false discovery: analyses of bibliometric indicators from large data sets. *Scientometrics*, on-line, 18 June 2013. Disponível em: <<http://link-springer->

com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-1063-z.pdf>. Acesso em: 10 ago. 2013.

RADICCHI, F.; CASTELLANO, C. Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics*, on-line, 30 Apr. 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-1027-3.pdf>>. Acesso em 10 ago. 2013.

SANTANA, O. A.; PEIXOTO, L. R. T. O impacto do Portal Capes nas referências de artigos científicos sobre Ciências Biológicas e Saúde na Universidade de Brasília. *RBPG*, Brasília, v. 7, n. 13, p. 352-62, nov. 2010. Disponível em: <http://www2.capes.gov.br/rbpg/images/stories/downloads/RBPG/Vol.7_13/6_Artigo.pdf>. Acesso em: 10 fev. 2011.

SANTOS, R. N. M.; KOBASHI, N. Y. Bibliometria, Cientometria, Infometria: conceitos e aplicações. *Pesq. bras. Ci. Inf.*, Brasília, v.2, n.1, p.155-172, jan./dez. 2009. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/viewArticle/21>>. Acesso em: 05 ago. 2013.

SCHUBERT, A.; BRAUN, T. International collaboration in the sciences, 1981-1985. *Scientometrics*, v. 19, n. 1-2, p. 3–10, 1990. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2FBF02130461.pdf>>. Acesso em: 10 ago. 2013.

SMALL H., GARFIELD E. The geography of science: disciplinary and national mappings. *Journal of Information Science*, [S. l.], v. 11, n. 4, p. 147-159, 1986. Disponível em: <<http://www.garfield.library.upenn.edu/essays/v9p324y1986.pdf>>. Acesso em 05 fev. 2013.

SOLANO, R. R. C.; VALDIVIA, A. M. El SIDA y su productividad científica en la base de datos SciELO entre 1997 - 2003: estudio bibliométrico. *Biblios [online]*, v. 4, n. 16, p. 81-92, dez. 2003. Disponível em: <http://www.bibliosperu.com/articulos/16/2003_030.pdf>. Acesso em: 12 ago. 2013.

SOORYAMOORTHY, R. Collaboration and publication: How collaborative are scientists in South Africa? *Scientometrics*, v. 80, n. 2, p. 419–439, 2009. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-008-2074-z.pdf>>. Acesso em: 10 ago. 2013.

SOTUDEH, H.; KHOSHIAN, N. Gender differences in science: the case of scientific productivity in Nano Science & Technology during 2005–2007. *Scientometrics*, on-line, 17 May 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-1031-7.pdf>>. Acesso em 10 ago. 2013.

SULEIMENOV, E. Z. et al. Creating the Kazakh Science Citation Index. *Scientific and Technical Information Processing*, v. 36, n. 3, p. 151–155, 2009. Disponível em: <<http://link->

springer-
com.ez27.periodicos.capes.gov.br/content/pdf/10.3103%2FS0147688209030046.pdf>.
Acesso em: 05 ago. 2013.

TAGUE-SUTCLIFFE, J. An introduction to informetrics. *Information processing & management*, Oxford, v. 28, n. 1, p. 1-3, 1992. Disponível em:
<<http://www.sciencedirect.com/science/article/pii/030645739290087G>>. Acesso em: 18 fev. 2011.

TRIPODI, T. *et al.* *Análise da pesquisa social*. Rio de Janeiro: Francisco Alves, 1975. 337 p.

VANTI, N. A cientometria revisitada à luz da expansão da ciência, da tecnologia e da inovação. *Ponto de Acesso*, Salvador, v. 5, p. 5-31, 2011. Disponível em:
<<http://repositorio.ufrn.br:8080/jspui/handle/1/6185>>. Acesso em: 10 ago. 2013.

VASCONCELOS, L. G. A RBCCV em versão eletrônica no SciELO. *Revista Brasileira de Cirurgia Cardiovascular [online]*, v. 14, n. 1, 1999. Disponível em:
<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-76381999000100002&lng=en&nrm=iso>. Acesso em: 12 ago. 2013.

VICENTINI, L. A. Gestão em bibliotecas digitais. In: MARCONDES, C. H. et al. (orgs.). *Bibliotecas digitais: saberes e práticas*. Salvador, BA: EDUFBA; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2 ed, 2006. 337 p.

VINE, R. Google Scholar. *Journal of the Medical Library Association*, v. 94, n. 1, p. 97–99, 2006. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1324783/>>. Acesso em: 10 ago. 2013.

WEBSTER, B. M. Polish sociology citation index as an example of usage of national citation indexes in scientometric analysis of social sciences. *Journal of Information Science*, v. 24, n.1, p. 19-32, 1998. Disponível em: <<http://jis.sagepub.com/content/24/1/19.abstract>>. Acesso em: 20 jan. 2013.

WINTER, J. C. F. de; ZADPOOR, A. A.; DODOU, D. The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics*, on-line, 01 Aug. 2013. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/article/10.1007/s11192-013-1089-2/fulltext.html>>. Acesso em: 10 ago. 2013.

WLEKLINSKI, J. M. Studying Google Scholar: Wall to wall coverage? *Online*, v. 29, n. 3, p. 22–26, 2005. Disponível em: <<http://connection.ebscohost.com/c/articles/16833207/studying-google-scholar-wall-wall-coverage>>. Acesso em: 10 ago. 2013.

XIN-NING, S. et al. Developing the Chinese social science citation index. *Online Information Review*, v. 25, n6, p. 365-69, 2001. Disponível em: < http://www.e-library.lt/resursai/Uzsienio%20leidiniai/Elibrary_ru/Index_nc/Publikacii/Xin-ning2001.pdf >. Acesso em: 20 jan. 2013.

YE, J. Development, significance and background information about the “Chinese Book Citation Index” (CBkCI) demonstration database. *Scientometrics*, on-line, 23 July 2013. Disponível em <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11192-013-1092-7.pdf>>. Acesso em 05 ago. 2013.

ZIBAREVA, I. V.; PARMON, V. N. Ranking Institutes of the Russian Academy of Sciences by the Russian Science Citation Index: The Case of Chemical Research Institutes. *Herald of the Russian Academy of Sciences*, v. 82, n. 5, p. 363-372, 2012. Disponível em: <<http://link-springer-com.ez27.periodicos.capes.gov.br/content/pdf/10.1134%2FS1019331612050048.pdf>>. Acesso em 05 ago. 2013.

ANEXO A – CONSULTAS AOS DADOS FONTE DO SciELO

FIGURA 55 – Lista de dados fonte na
Coleção Saúde Pública: ISSN 0036-3634



Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Salud pública Méx	127	1279	43387	2875	10.07
2013	2	18	810	69	9.00
2012	7	70	2529	219	10.00
2011	10	101	3769	246	10.10
2010	8	93	2963	276	11.62
2009	10	119	5088	333	11.90
2008	10	104	3963	285	10.40
2007	10	97	3117	244	9.70
2006	8	85	2832	227	10.62
2005	6	46	1472	145	7.67
2004	6	45	1496	124	7.50
2003	11	120	3802	186	10.91
2002	7	78	2409	78	11.14
2001	6	61	1933	67	10.17
2000	6	52	1252	68	8.67
1999	8	68	2106	123	8.50
1998	6	55	1694	101	9.17
1997	6	67	2152	84	11.17
total	127	1279	43387	2875	10.07

Fonte: SciELO, 2013⁵⁷

FIGURA 56 – Lista de dados fonte na
Coleção México: ISSN 0036-3634



Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Salud pública Méx	98	1014	35064	4761	10.35
2012	6	61	2217	567	10.17
2011	10	101	3769	576	10.10
2010	8	93	2963	597	11.62
2009	10	119	5088	624	11.90
2008	10	108	3966	593	10.80
2007	10	97	3117	416	9.70
2006	8	85	2832	426	10.62
2005	6	46	1472	266	7.67
2004	6	45	1496	263	7.50
2003	11	120	3802	229	10.91
2002	7	78	2409	110	11.14
2001	6	61	1933	94	10.17
total	98	1014	35064	4761	10.35

Fonte: SciELO, 2013⁵⁸

⁵⁷ Disponível em: <[http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=0036-3634&CITED\[\]=SALUD+PUBLICA+DE+MEXICO&YNG\[\]=all](http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=0036-3634&CITED[]=SALUD+PUBLICA+DE+MEXICO&YNG[]=all)>. Acesso em: 13 jun. 2013.

⁵⁸ Disponível em: <[http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=mex&issn=0036-3634&CITED\[\]=SALUD+PUBLICA+DE+MEXICO&YNG\[\]=all](http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=mex&issn=0036-3634&CITED[]=SALUD+PUBLICA+DE+MEXICO&YNG[]=all)>. Acesso em: 13 jun. 2013.

FIGURA 57 – Lista de dados fonte na
Coleção Saúde Pública: ISSN 0124-0064



Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev. salud pública	47	575	13949	302	12.23
2012	4	50	1228	31	12.50
2011	6	87	2099	47	14.50
2010	8	100	2625	67	12.50
2009	6	89	2208	42	14.83
2008	6	88	2190	47	14.67
2007	4	54	1172	19	13.50
2006	5	52	1003	21	10.40
2005	3	28	646	13	9.33
2004	4	22	660	11	5.50
2003	1	5	118	4	5.00
total	47	575	13949	302	12.23

Fonte: SciELO, 2013⁵⁹

FIGURA 58 – Lista de dados fonte na
Coleção Colômbia: ISSN 0124-0064



Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev. salud pública	57	644	15759	1135	11.30
2012	5	64	1546	240	12.80
2011	6	87	2115	239	14.50
2010	8	100	2605	202	12.50
2009	6	89	2208	145	14.83
2008	6	88	2190	114	14.67
2007	4	54	1168	61	13.50
2006	5	52	1246	53	10.40
2005	3	28	646	25	9.33
2004	4	22	660	16	5.50
2003	3	18	482	8	6.00
2002	4	25	427	20	6.25
2001	3	17	466	12	5.67
total	57	644	15759	1135	11.30

Fonte: SciELO, 2013⁶⁰

⁵⁹ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=0124-0064&CITED%5B%5D=REVISTA+DE+SALUD+PUBLICA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

⁶⁰ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=col&issn=0124-0064&CITED%5B%5D=REVISTA+DE+SALUD+PUBLICA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

FIGURA 59 – Lista de dados fonte na
Coleção Saúde Pública: ISSN 0213-9111



Gaceta Sanitaria
ISSN 0213-9111

Data do último
processamento
11-06-2013

Lista de dados fonte

▼ - clique para seleccionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Gac. sanit. (Barc.)	68	826	22552	2019	12.15
2012	6	85	2751	281	14.17
2011	6	80	2638	298	13.33
2010	6	78	2383	293	13.00
2009	6	93	3136	304	15.50
2008	6	79	2339	250	13.17
2007	6	67	1694	137	11.17
2006	6	63	128	20	10.50
2005	6	58	1527	133	9.67
2004	7	93	2609	95	13.29
2003	7	72	1838	110	10.29
2002	6	58	1509	98	9.67
total	68	826	22552	2019	12.15

Fonte: SciELO, 2013⁶¹

FIGURA 60 – Lista de dados fonte na
Coleção Espanha: ISSN 0213-9111



Gaceta Sanitaria
ISSN 0213-9111

Data do último
processamento
11-06-2013

Lista de dados fonte

▼ - clique para seleccionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Gac. sanit. (Barc.)	71	866	24963	3411	12.20
2013	2	30	855	109	15.00
2012	6	85	2751	490	14.17
2011	6	80	2678	516	13.33
2010	6	78	2383	453	13.00
2009	6	93	3136	447	15.50
2008	6	79	2339	352	13.17
2007	6	67	1639	182	11.17
2006	6	63	1521	267	10.50
2005	6	58	1527	172	9.67
2004	8	102	2784	188	12.75
2003	7	73	1843	128	10.43
2002	6	58	1507	107	9.67
total	71	866	24963	3411	12.20

Fonte: SciELO, 2013⁶²

⁶¹ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=0213-9111&CITED%5B%5D=GACETA+SANITARIA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

⁶² Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=esp&issn=0213-9111&CITED%5B%5D=GACETA+SANITARIA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

FIGURA 61 – Lista de dados fonte na
Coleção Saúde Pública: ISSN 0864-3466

<p>REVISTA CUBANA DE SALUD PUBLICA</p> <p>Data do último processamento 11-06-2013</p>	<p>Revista Cubana de Salud Pública ISSN 0864-3466</p>
--	--

Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev Cubana Salud Pública	25	334	7147	347	13.36
2012	5	64	1507	96	12.80
2011	4	55	1067	73	13.75
2010	4	56	1278	59	14.00
2009	4	51	1105	51	12.75
2008	4	53	921	39	13.25
2007	4	55	1269	29	13.75
total	25	334	7147	347	13.36

Fonte: SciELO, 2013⁶³

FIGURA 62 – Lista de dados fonte na
Coleção de Cuba: ISSN 0864-3466

<p>REVISTA CUBANA DE SALUD PUBLICA</p> <p>Data do último processamento 11-06-2013</p>	<p>Revista Cubana de Salud Pública ISSN 0864-3466</p>
--	--

Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev Cubana Salud Pública	60	666	12279	1610	11.10
2013	2	31	843	109	15.50
2012	5	64	1507	325	12.80
2011	5	55	1456	327	11.00
2010	4	56	1278	239	14.00
2009	4	51	954	185	12.75
2008	4	59	871	118	14.75
2007	4	40	868	77	10.00
2006	4	56	548	46	14.00
2005	4	34	695	15	8.50
2004	4	42	585	36	10.50
2003	4	46	954	18	11.50
2002	3	24	390	38	8.00
2001	2	16	241	15	8.00
2000	2	19	175	14	9.50
1999	2	18	237	15	9.00
1998	2	16	160	11	8.00
1997	1	9	117	12	9.00
1996	2	16	205	5	8.00
1995	2	14	195	5	7.00
total	60	666	12279	1610	11.10

Fonte: SciELO, 2013⁶⁴

⁶³ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=0864-3466&CITED%5B%5D=revista%20cubana%20de%20salud%20publica&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

⁶⁴ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=cub&issn=0864-3466&CITED%5B%5D=revista%20cubana%20de%20salud%20publica&YNG%5B%5D=all>. Acesso em: 13 jun. 2013

FIGURA 63 – Lista de dados fonte na
Coleção Saúde Pública: ISSN 1135-5727



▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev. esp. salud pública	96	805	23697	971	8.39
2012	6	50	1843	93	8.33
2011	6	55	1757	92	9.17
2010	6	60	1749	127	10.00
2009	6	66	2594	149	11.00
2008	6	49	1762	115	8.17
2007	6	46	1658	78	7.67
2006	6	46	46	18	7.67
2005	6	49	1545	46	8.17
2004	6	47	1455	33	7.83
2003	6	46	1395	54	7.67
2002	6	56	1712	47	9.33
2001	6	41	1118	28	6.83
2000	6	54	1447	22	9.00
1999	6	53	1262	52	8.83
1998	6	51	1203	15	8.50
1997	6	36	1151	2	6.00
total	96	805	23697	971	8.39

Fonte: SciELO, 2013⁶⁵

FIGURA 64 – Lista de dados fonte na
Coleção Espanha: ISSN 1135-5727



▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev. esp. salud pública	97	811	25235	2040	8.36
2013	1	6	280	54	6.00
2012	6	50	1843	270	8.33
2011	6	55	1757	294	9.17
2010	6	60	1749	278	10.00
2009	6	66	2594	270	11.00
2008	6	49	1747	207	8.17
2007	6	46	1557	136	7.67
2006	6	46	1400	109	7.67
2005	6	49	1545	91	8.17
2004	6	47	1455	66	7.83
2003	6	46	1415	83	7.67
2002	6	56	1712	52	9.33
2001	6	41	1118	38	6.83
2000	6	54	1447	24	9.00
1999	6	53	1262	51	8.83
1998	6	51	1203	15	8.50
1997	6	36	1151	2	6.00
total	97	811	25235	2040	8.36

Fonte: SciELO, 2013⁶⁶

⁶⁵ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=1135-5727&CITED%5B%5D=REVISTA+ESPANOLA+DE+SALUD+PUBLICA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

⁶⁶ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=esp&issn=1135-5727&CITED%5B%5D=REVISTA+ESPANOLA+DE+SALUD+PUBLICA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

FIGURA 65 – Lista de dados fonte na
Coleção Saúde Pública: ISSN 1413-8123

Ciência & Saúde Coletiva					
Data do último processamento 11-06-2013					
Ciência & Saúde Coletiva ISSN 1413-8123					
Lista de dados fonte					
▼ - clique para selecionar a coluna de ordenação ▲ - indica a ordem corrente					
titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Ciênc. saúde coletiva	90	2405	74338	2618	26.72
2013	5	144	4622	98	28.80
2012	12	308	10311	412	25.67
2011	13	457	14609	376	35.15
2010	9	356	10350	278	39.56
2009	7	212	6494	293	30.29
2008	8	211	6514	352	26.38
2007	7	162	4856	179	23.14
2006	5	109	2907	135	21.80
2005	5	118	3560	192	23.60
2004	4	85	2411	120	21.25
2003	4	76	2484	78	19.00
2002	4	64	1994	51	16.00
2001	2	33	793	31	16.50
2000	2	27	1063	9	13.50
1999	2	30	1060	9	15.00
1998	1	13	310	5	13.00
total	90	2405	74338	2618	26.72

Fonte: SciELO, 2013⁶⁷

FIGURA 66 – Lista de dados fonte na
Coleção Brasil: ISSN 1413-8123

Ciência & Saúde Coletiva					
Data do último processamento 11-06-2013					
Ciência & Saúde Coletiva ISSN 1413-8123					
Lista de dados fonte					
▼ - clique para selecionar a coluna de ordenação ▲ - indica a ordem corrente					
titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Ciênc. saúde coletiva	91	2435	75377	7951	26.76
2013	6	174	5661	497	29.00
2012	12	308	10311	1666	25.67
2011	13	457	14609	1412	35.15
2010	9	356	10350	1139	39.56
2009	7	212	6494	1034	30.29
2008	8	211	6514	761	26.38
2007	7	162	4856	428	23.14
2006	5	109	2907	284	21.80
2005	5	118	3560	299	23.60
2004	4	85	2411	175	21.25
2003	4	76	2484	107	19.00
2002	4	64	1994	64	16.00
2001	2	33	793	45	16.50
2000	2	27	1063	17	13.50
1999	2	30	1060	11	15.00
1998	1	13	310	12	13.00
total	91	2435	75377	7951	26.76

Fonte: SciELO, 2013⁶⁸

⁶⁷ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=1413-8123&CITED%5B%5D=CIENCIA+%26+SAUDE+COLETIVA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

⁶⁸ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=scl&issn=1413-8123&CITED%5B%5D=CIENCIA+%26+SAUDE+COLETIVA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

FIGURA 67 – Lista de dados fonte na
Coleção Saúde Pública: ISSN 0102-311X

GSP CADERNOS DE SAÚDE PÚBLICA
REPORTS IN PUBLIC HEALTH

Cadernos de Saúde Pública
ISSN 0102-311x

Data do último
processamento
11-06-2013

Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Cad. saúde pública	220	3991	116703	13581	18.14
2013	5	88	3158	618	17.60
2012	13	228	7423	1700	17.54
2011	15	264	8752	1955	17.60
2010	12	214	6872	1510	17.83
2009	15	301	9945	1468	20.07
2008	16	347	10974	1500	21.69
2007	16	337	10365	1088	21.06
2006	13	260	7845	892	20.00
2005	7	214	5923	730	30.57
2004	8	215	6180	487	26.88
2003	8	230	6077	372	28.75
2002	7	189	5728	298	27.00
2001	7	158	4765	203	22.57
2000	6	131	3776	154	21.83
1999	6	110	3083	143	18.33
1998	7	131	3082	136	18.71
1997	6	96	2312	76	16.00
1996	6	77	1778	74	12.83
1995	4	54	1253	41	13.50
1994	6	73	1857	31	12.17
1993	5	56	1336	29	11.20
1992	4	46	887	16	11.50
1991	4	38	869	20	9.50
1990	4	29	558	17	7.25
1989	4	18	269	6	4.50
1988	4	17	407	5	4.25
1987	4	25	445	2	6.25
1986	4	20	377	7	5.00
1985	4	25	407	3	6.25
total	220	3991	116703	13581	18.14

Fonte: SciELO, 2013⁶⁹

FIGURA 68 – Lista de dados fonte na
Coleção Brasil: ISSN 0102-311X

GSP CADERNOS DE SAÚDE PÚBLICA
REPORTS IN PUBLIC HEALTH

Cadernos de Saúde Pública
ISSN 0102-311x

Data do último
processamento
11-06-2013

Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Cad. saúde pública	220	3990	116679	28050	18.14
2013	5	88	3158	1348	17.60
2012	13	228	7423	4238	17.54
2011	15	264	8752	4194	17.60
2010	12	214	6872	3723	17.83
2009	15	301	9945	3314	20.07
2008	16	347	10974	2978	21.69
2007	16	338	10405	2144	21.12
2006	13	260	7845	1713	20.00
2005	7	214	5923	1244	30.57
2004	8	215	6180	808	26.88
2003	8	229	6013	587	28.62
2002	7	189	5728	443	27.00
2001	7	158	4765	335	22.57
2000	6	131	3776	214	21.83
1999	6	110	3083	192	18.33
1998	7	131	3082	178	18.71
1997	6	96	2312	100	16.00
1996	6	77	1778	96	12.83
1995	4	54	1253	49	13.50
1994	6	73	1857	35	12.17
1993	5	56	1336	31	11.20
1992	4	46	887	19	11.50
1991	4	38	869	22	9.50
1990	4	29	558	18	7.25
1989	4	18	269	9	4.50
1988	4	17	407	5	4.25
1987	4	24	445	3	6.00
1986	4	20	377	7	5.00
1985	4	25	407	3	6.25
total	220	3990	116679	28050	18.14

Fonte: SciELO, 2013⁷⁰

⁶⁹ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=0102-311x&CITED%5B%5D=CADERNOS+DE+SAUDE+PUBLICA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

⁷⁰ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=scl&issn=0102-311x&CITED%5B%5D=CADERNOS+DE+SAUDE+PUBLICA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

FIGURA 69 – Lista de dados fonte na Coleção Saúde Pública: ISSN 0034-8910

RSP Revista de Saúde Pública

Revista de Saúde Pública
ISSN 0034-8910

Data do último processamento
11-06-2013

Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev. saúde pública	263	3511	67720	11756	13.35
2013	2	25	567	387	12.50
2012	7	138	3039	1148	19.71
2011	6	137	3168	1262	22.83
2010	6	132	2764	975	22.00
2009	8	157	3056	945	19.62
2008	8	170	3257	1063	21.25
2007	8	157	3048	788	19.62
2006	8	161	3285	564	20.12
2005	6	131	2378	518	21.83
2004	7	129	2254	465	18.43
2003	6	115	2072	369	19.17
2002	7	121	2360	246	17.29
2001	6	87	1325	250	14.50
2000	7	112	1874	261	16.00
1999	6	79	1479	166	13.17
1998	6	85	1883	170	14.17
1997	7	88	2080	231	12.57
1996	6	78	1929	163	13.00
1995	6	69	1573	143	11.50
1994	6	65	1599	121	10.83
1993	7	65	1592	137	9.29
1992	6	64	1346	84	10.67
1991	6	68	1578	89	11.33
1990	7	67	1629	91	9.57
1989	6	62	1187	88	10.33
1988	6	61	1311	91	10.17
1987	6	51	1100	80	8.50
1986	6	53	865	64	8.83
1985	6	56	976	68	9.33
1984	7	53	1020	67	7.57
1983	6	40	1084	93	6.67
1982	7	34	664	60	4.86
1981	7	64	1239	85	9.14
1980	4	52	1108	83	13.00
1979	6	44	688	45	7.33
1978	4	51	836	60	12.75
1977	4	54	734	50	13.50
1976	6	40	605	24	6.67
1975	4	55	777	53	13.75
1974	5	45	513	18	9.00
1973	4	41	435	37	10.25
1972	4	43	310	14	10.75
1971	2	28	294	15	14.00
1970	2	24	178	17	12.00
1969	2	26	199	7	13.00
1968	2	13	261	0	6.50
1967	2	21	201	1	10.50
total	263	3511	67720	11756	13.35

Fonte: SciELO, 2013⁷¹

⁷¹ Disponível em:
<http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=0034-8910&CITED%5B%5D=REVISTA+DE+SAUDE+PUBLICA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

FIGURA 70 – Lista de dados fonte na Coleção Brasil: ISSN 0034-8910

Revista de Saúde Pública	Revista de Saúde Pública				
ISSN 0034-8910	ISSN 0034-8910				
<small>Data do último processamento 11-06-2013</small>	Lista de dados fonte				
▽ - clique para selecionar a coluna de ordenação					
△ - indica a ordem corrente					
titulo da revista/ano △	n. de fascículos ▽	n. de artigos ▽	n. de citações concedidas ▽	n. de citações recebidas ▽	média de artigos por fascículo ▽
Rev. saúde pública	263	3511	67724	25292	13.35
2013	2	25	567	963	12.50
2012	7	138	3043	3147	19.71
2011	6	137	3168	3064	22.83
2010	6	132	2764	2689	22.00
2009	8	157	3056	2355	19.62
2008	8	170	3257	2361	21.25
2007	8	157	3048	1781	19.62
2006	8	161	3285	1392	20.12
2005	6	131	2378	1108	21.83
2004	7	129	2254	921	18.43
2003	6	115	2072	710	19.17
2002	7	121	2360	586	17.29
2001	6	87	1325	517	14.50
2000	7	112	1874	430	16.00
1999	6	79	1479	314	13.17
1998	6	85	1883	306	14.17
1997	7	88	2080	303	12.57
1996	6	78	1929	256	13.00
1995	6	69	1573	185	11.50
1994	6	65	1599	167	10.83
1993	7	65	1592	162	9.29
1992	6	64	1346	122	10.67
1991	6	68	1578	112	11.33
1990	7	67	1629	120	9.57
1989	6	62	1187	116	10.33
1988	6	61	1311	117	10.17
1987	6	51	1100	89	8.50
1986	6	53	865	81	8.83
1985	6	56	976	81	9.33
1984	7	53	1020	73	7.57
1983	6	40	1084	94	6.67
1982	7	34	664	60	4.86
1981	7	64	1239	85	9.14
1980	4	52	1108	84	13.00
1979	6	44	688	45	7.33
1978	4	51	836	60	12.75
1977	4	54	734	50	13.50
1976	6	40	605	24	6.67
1975	4	55	777	53	13.75
1974	5	45	513	18	9.00
1973	4	41	435	37	10.25
1972	4	43	310	14	10.75
1971	2	28	294	15	14.00
1970	2	24	178	17	12.00
1969	2	26	199	7	13.00
1968	2	13	261	0	6.50
1967	2	21	201	1	10.50
total	263	3511	67724	25292	13.35

Fonte: SciELO, 2013⁷²

⁷² Disponível em:
http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=scl&issn=0034-8910&CITED%5B%5D=REVISTA+DE+SAUDE+PUBLICA&YNG%5B%5D=all. Acesso em: 13 jun. 2013.

FIGURA 71 – Lista de dados fonte na
Coleção Saúde Pública: ISSN 1415-790x

revista brasileira de
epidemiologia

Revista Brasileira de Epidemiologia
ISSN 1415-790x

Data do último
processamento
11-06-2013

Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev. bras. epidemiol	57	672	19600	1392	11.79
2013	1	21	603	100	21.00
2012	4	79	2499	259	19.75
2011	5	80	2300	261	16.00
2010	4	62	2025	206	15.50
2009	4	61	1875	160	15.25
2008	5	73	2297	147	14.60
2007	4	58	1568	100	14.50
2006	4	41	1168	59	10.25
2005	5	42	1016	27	8.40
2004	4	38	915	24	9.50
2003	4	33	706	17	8.25
2002	4	33	951	9	8.25
2001	3	15	521	12	5.00
2000	1	5	161	7	5.00
1999	2	9	208	3	4.50
1998	3	22	787	1	7.33
total	57	672	19600	1392	11.79

Fonte: SciELO, 2013⁷³

FIGURA 72 – Lista de dados fonte na
Coleção Brasil: ISSN 1415-790x

revista brasileira de
epidemiologia

Revista Brasileira de Epidemiologia
ISSN 1415-790x

Data do último
processamento
11-06-2013

Lista de dados fonte

▼ - clique para selecionar a coluna de ordenação
▲ - indica a ordem corrente

titulo da revista/ano ▲	n. de fascículos ▼	n. de artigos ▼	n. de citações concedidas ▼	n. de citações recebidas ▼	média de artigos por fascículo ▼
Rev. bras. epidemiol	57	672	19600	2861	11.79
2013	1	21	603	169	21.00
2012	4	79	2499	598	19.75
2011	5	80	2300	569	16.00
2010	4	62	2025	485	15.50
2009	4	61	1875	357	15.25
2008	5	73	2297	277	14.60
2007	4	58	1568	167	14.50
2006	4	41	1168	97	10.25
2005	5	42	1016	41	8.40
2004	4	38	915	38	9.50
2003	4	33	706	23	8.25
2002	4	33	951	13	8.25
2001	3	15	521	13	5.00
2000	1	5	161	9	5.00
1999	2	9	208	4	4.50
1998	3	22	787	1	7.33
total	57	672	19600	2861	11.79

Fonte: SciELO, 2013⁷⁴

⁷³ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=spa&issn=1415-790x&CITED%5B%5D=REVISTA+BRASILEIRA+DE+EPIDEMIOLOGIA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

⁷⁴ Disponível em: <http://statbiblio.scielo.org/stat_biblio/index.php?state=15&lang=pt&country=scl&issn=1415-790x&CITED%5B%5D=REVISTA+BRASILEIRA+DE+EPIDEMIOLOGIA&YNG%5B%5D=all>. Acesso em: 13 jun. 2013.

ANEXO B – CONSULTAS AOS DADOS DA BASE DE CITAÇÕES⁷⁵

FIGURA 73 – Dados comparativos da Coleção Saúde Pública: ISSN 1851-8265

1851-8265	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	26	26	0	200	165	35 (21,21%)	7,69	6,35	1,34 (21,11%)	5.820	5.833	-13 (-0,22%)
2012	4	4	0	<u>38</u>	26	12 (46,15%)	9,50	6,50	3,00 (46,15%)	1.002	1.002	0
2011	4	4	0	<u>41</u>	26	15 (57,69%)	10,25	6,50	3,75 (57,69%)	972	985	-13 (-1,32%)
2010	3	3	0	<u>24</u>	20	4 (20,00%)	8,00	6,67	1,33 (19,94%)	658	658	0
2009	3	3	0	<u>28</u>	21	7 (33,33%)	9,33	7,00	2,33 (33,29%)	902	902	0
2008	3	3	0	<u>27</u>	18	9 (50,00%)	9,00	6,00	3,00 (50,00%)	844	844	0
2007	3	3	0	<u>28</u>	20	8 (40,00%)	9,33	6,67	2,66 (39,88%)	339	339	0
2006	3	3	0	<u>24</u>	19	5 (26,32%)	8,00	6,33	1,67 (26,38%)	564	564	0
2005	3	3	0	<u>18</u>	15	3 (20,00%)	6,00	5,00	1,00 (20,00%)	539	539	0

Fonte: consulta ao banco de dados⁷⁶

⁷⁵ Em relação à coluna de total de artigos da BC: a linha TOTAL representa o total de artigos que possuem no mínimo uma citação; as linhas anuais mostram o total de artigos para cada ano, independente de possuírem ou não citações associadas a ele. Nesse caso, a soma do número de artigos em todos os anos não corresponderá ao total de artigos na BC. Essa observação é válida para todas as FIGURAS deste Anexo.

⁷⁶ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 74 – Dados comparativos da Coleção Saúde Pública: ISSN 1726-4634

1726-4634	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	13	13	0	349	302	47 (15,56%)	26,85	23,23	3,62 (15,58%)	7.104	7.104	0
2013	1	1	0	<u>39</u>	24	15 (62,50%)	39,00	24,00	15,00 (62,50%)	595	595	0
2012	4	4	0	<u>114</u>	100	14 (14,00%)	28,50	25,00	3,50 (14,00%)	2.025	2.025	0
2011	4	4	0	<u>124</u>	102	22 (21,57%)	31,00	25,50	5,50 (21,57%)	2.334	2.334	0
2010	4	4	0	<u>112</u>	76	36 (47,37%)	28,00	19,00	9,00 (47,37%)	2.150	2.150	0

Fonte: consulta ao banco de dados⁷⁷

FIGURA 75 – Dados comparativos da Coleção Saúde Pública: ISSN 1555-7960

1555-7960	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	7	7	0	65	52	13 (25,00%)	9,29	7,43	1,86 (25,04%)	1.781	1.781	0
2013	1	1	0	<u>10</u>	6	4 (66,67%)	10,00	6,00	4,00 (66,67%)	221	221	0
2012	4	4	0	<u>50</u>	30	20 (66,67%)	12,50	7,50	5,00 (66,67%)	1.049	1.049	0
2011	2	2	0	<u>28</u>	16	12 (75,00%)	14,00	8,00	6,00 (75,00%)	511	511	0

Fonte: consulta ao banco de dados⁷⁸

⁷⁷ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

⁷⁸ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 76 – Dados comparativos da Coleção Saúde Pública: ISSN 1415-790x

1415-790x	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	56	57	-1 (-1,75%)	670	671	-1 (-0,15%)	11,96	11,77	0,19 (1,61%)	19.107	19.567	-460 (-2,35%)
2013	1	1	0	4	20	-16 (-80,00%)	4,00	20,00	-16,00 (-80,00%)	124	570	-446 (-78,25%)
2012	4	4	0	83	79	4 (5,06%)	20,75	19,75	1,00 (5,06%)	2.499	2.499	0
2011	5	5	0	84	80	4 (5,00%)	16,80	16,00	0,80 (5,00%)	2.300	2.300	0
2010	4	4	0	65	62	3 (4,84%)	16,25	15,50	0,75 (4,84%)	2.025	2.025	0
2009	4	4	0	64	61	3 (4,92%)	16,00	15,25	0,75 (4,92%)	1.875	1.875	0
2008	5	5	0	92	73	19 (26,03%)	18,40	14,60	3,80 (26,03%)	2.297	2.297	0
2007	4	4	0	66	58	8 (13,79%)	16,50	14,50	2,00 (13,79%)	1.568	1.568	0
2006	4	4	0	61	41	20 (48,78%)	15,25	10,25	5,00 (48,78%)	1.168	1.168	0
2005	5	5	0	58	42	16 (38,10%)	11,60	8,40	3,20 (38,10%)	1.002	1.016	-14 (-1,38%)
2004	4	4	0	47	38	9 (23,68%)	11,75	9,50	2,25 (23,68%)	915	915	0
2003	4	4	0	44	33	11 (33,33%)	11,00	8,25	2,75 (33,33%)	706	706	0
2002	4	4	0	41	33	8 (24,24%)	10,25	8,25	2,00 (24,24%)	951	951	0
2001	3	3	0	18	15	3 (20,00%)	6,00	5,00	1,00 (20,00%)	521	521	0
2000	1	1	0	7	5	2 (40,00%)	7,00	5,00	2,00 (40,00%)	161	161	0
1999	2	2	0	11	9	2 (22,22%)	5,50	4,50	1,00 (22,22%)	208	208	0
1998	3	3	0	27	22	5 (22,73%)	9,00	7,33	1,67 (22,78%)	787	787	0

Fonte: consulta ao banco de dados⁷⁹

⁷⁹ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 77 – Dados comparativos da Coleção Saúde Pública: ISSN 1413-8123

1413-8123	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	87	90	-3 (-3,33%)	2.546	2.405	141 (5,86%)	29,26	26,72	2,54 (9,51%)	74.338	74.338	0
2013	5	5	0	165	144	21 (14,58%)	33,00	28,80	4,20 (14,58%)	4.622	4.622	0
2012	12	12	0	375	308	67 (21,75%)	31,25	25,67	5,58 (21,74%)	10.311	10.311	0
2011	13	13	0	522	457	65 (14,22%)	40,15	35,15	5,00 (14,22%)	14.609	14.609	0
2010	7	9	-2 (-22,22%)	401	356	45 (12,64%)	57,29	39,56	17,73 (44,82%)	10.350	10.350	0
2009	7	7	0	267	212	55 (25,94%)	38,14	30,29	7,85 (25,92%)	6.494	6.494	0
2008	7	8	-1 (-12,50%)	268	211	57 (27,01%)	38,29	26,38	11,91 (45,15%)	6.514	6.514	0
2007	7	7	0	207	162	45 (27,78%)	29,57	23,14	6,43 (27,79%)	4.856	4.856	0
2006	5	5	0	149	109	40 (36,70%)	29,80	21,80	8,00 (36,70%)	2.907	2.907	0
2005	5	5	0	165	118	47 (39,83%)	33,00	23,60	9,40 (39,83%)	3.560	3.560	0
2004	4	4	0	112	85	27 (31,76%)	28,00	21,25	6,75 (31,76%)	2.411	2.411	0
2003	4	4	0	98	76	22 (28,95%)	24,50	19,00	5,50 (28,95%)	2.484	2.484	0
2002	4	4	0	77	64	13 (20,31%)	19,25	16,00	3,25 (20,31%)	1.994	1.994	0
2001	2	2	0	40	33	7 (21,21%)	20,00	16,50	3,50 (21,21%)	793	793	0
2000	2	2	0	35	27	8 (29,63%)	17,50	13,50	4,00 (29,63%)	1.063	1.063	0
1999	2	2	0	37	30	7 (23,33%)	18,50	15,00	3,50 (23,33%)	1.060	1.060	0
1998	1	1	0	16	13	3 (23,08%)	16,00	13,00	3,00 (23,08%)	310	310	0

Fonte: consulta ao banco de dados⁸⁰

⁸⁰ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 78 – Dados comparativos da Coleção Saúde Pública: ISSN 1135-5727

1135-5727	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	96	96	0	835	805	30 (3,73%)	8,70	8,39	0,31 (3,70%)	24.944	24.955	-11 (-0,04%)
2012	6	6	0	58	50	8 (16,00%)	9,67	8,33	1,34 (16,09%)	1.843	1.843	0
2011	6	6	0	63	55	8 (14,55%)	10,50	9,17	1,33 (14,50%)	1.757	1.757	0
2010	6	6	0	72	60	12 (20,00%)	12,00	10,00	2,00 (20,00%)	1.749	1.749	0
2009	6	6	0	80	66	14 (21,21%)	13,33	11,00	2,33 (21,18%)	2.594	2.594	0
2008	6	6	0	61	49	12 (24,49%)	10,17	8,17	2,00 (24,48%)	1.747	1.747	0
2007	6	6	0	64	46	18 (39,13%)	10,67	7,67	3,00 (39,11%)	1.557	1.557	0
2006	6	6	0	68	46	22 (47,83%)	11,33	7,67	3,66 (47,72%)	1.389	1.400	-11 (-0,79%)
2005	6	6	0	64	49	15 (30,61%)	10,67	8,17	2,50 (30,60%)	1.545	1.545	0
2004	6	6	0	67	47	20 (42,55%)	11,17	7,83	3,34 (42,66%)	1.455	1.455	0
2003	6	6	0	68	46	22 (47,83%)	11,33	7,67	3,66 (47,72%)	1.415	1.415	0
2002	6	6	0	67	56	11 (19,64%)	11,17	9,33	1,84 (19,72%)	1.712	1.712	0
2001	6	6	0	59	41	18 (43,90%)	9,83	6,83	3,00 (43,92%)	1.118	1.118	0
2000	6	6	0	68	54	14 (25,93%)	11,33	9,00	2,33 (25,89%)	1.447	1.447	0
1999	6	6	0	74	53	21 (39,62%)	12,33	8,83	3,50 (39,64%)	1.262	1.262	0
1998	6	6	0	68	51	17 (33,33%)	11,33	8,50	2,83 (33,29%)	1.203	1.203	0
1997	6	6	0	49	36	13 (36,11%)	8,17	6,00	2,17 (36,17%)	1.151	1.151	0

Fonte: consulta ao banco de dados⁸¹

⁸¹ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 79 – Dados comparativos da Coleção Saúde Pública: ISSN 1020-4989

1020-4989	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	186	186	0	1.693	1.769	-76 (-4,30%)	9,10	9,51	-0,41 (-4,31%)	49.274	49.858	-584 (-1,17%)
2013	2	2	0	21	20	1 (5,00%)	10,50	10,00	0,50 (5,00%)	578	578	0
2012	12	12	0	148	138	10 (7,25%)	12,33	11,50	0,83 (7,22%)	4.624	4.624	0
2011	12	12	0	164	148	16 (10,81%)	13,67	12,33	1,34 (10,87%)	4.817	4.817	0
2010	12	12	0	131	121	10 (8,26%)	10,92	10,08	0,84 (8,33%)	4.289	4.289	0
2009	12	12	0	154	145	9 (6,21%)	12,83	12,08	0,75 (6,21%)	4.965	4.965	0
2008	12	12	0	140	100	40 (40,00%)	11,67	8,33	3,34 (40,10%)	3.450	3.450	0
2007	11	11	0	139	91	48 (52,75%)	12,64	8,27	4,37 (52,84%)	2.986	2.986	0
2006	11	11	0	160	89	71 (79,78%)	14,55	8,09	6,46 (79,85%)	3.026	3.026	0
2005	10	10	0	141	97	44 (45,36%)	14,10	9,70	4,40 (45,36%)	2.882	2.882	0
2004	12	12	0	168	100	68 (68,00%)	14,00	8,33	5,67 (68,07%)	2.737	2.737	0
2003	11	11	0	182	104	78 (75,00%)	16,55	9,45	7,10 (75,13%)	2.348	2.393	-45 (-1,88%)
2002	11	11	0	196	106	90 (84,91%)	17,82	9,64	8,18 (84,85%)	2.127	2.219	-92 (-4,15%)
2001	12	12	0	188	97	91 (93,81%)	15,67	8,08	7,59 (93,94%)	1.817	1.817	0
2000	11	11	0	183	97	86 (88,66%)	16,64	8,82	7,82 (88,66%)	2.117	2.168	-51 (-2,35%)
1999	11	11	0	177	96	81 (84,38%)	16,09	8,73	7,36 (84,31%)	1.861	2.054	-193 (-9,40%)
1998	12	12	0	211	108	103 (95,37%)	17,58	9,00	8,58 (95,33%)	2.326	2.375	-49 (-2,06%)
1997	12	12	0	209	112	97 (86,61%)	17,42	9,33	8,09 (86,71%)	2.324	2.478	-154 (-6,21%)

Fonte: consulta ao banco de dados⁸²

⁸² Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 80 – Dados comparativos da Coleção Saúde Pública: ISSN 0864-3466

0864-3466	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	57	60	-3 (-5,00%)	612	666	-54 (-8,11%)	10,74	11,10	-0,36 (-3,24%)	11.886	12.279	-393 (-3,20%)
2013	2	2	0	21	31	-10 (-32,26%)	10,50	15,50	-5,00 (-32,26%)	450	843	-393 (-46,62%)
2012	5	5	0	82	64	18 (28,13%)	16,40	12,80	3,60 (28,13%)	1.507	1.507	0
2011	5	5	0	86	55	31 (56,36%)	17,20	11,00	6,20 (56,36%)	1.456	1.456	0
2010	4	4	0	68	56	12 (21,43%)	17,00	14,00	3,00 (21,43%)	1.278	1.278	0
2009	4	4	0	72	51	21 (41,18%)	18,00	12,75	5,25 (41,18%)	954	954	0
2008	4	4	0	75	59	16 (27,12%)	18,75	14,75	4,00 (27,12%)	871	871	0
2007	4	4	0	74	40	34 (85,00%)	18,50	10,00	8,50 (85,00%)	868	868	0
2006	4	4	0	69	56	13 (23,21%)	17,25	14,00	3,25 (23,21%)	548	548	0
2005	4	4	0	49	34	15 (44,12%)	12,25	8,50	3,75 (44,12%)	695	695	0
2004	4	4	0	64	42	22 (52,38%)	16,00	10,50	5,50 (52,38%)	585	585	0
2003	4	4	0	59	46	13 (28,26%)	14,75	11,50	3,25 (28,26%)	954	954	0
2002	3	3	0	32	24	8 (33,33%)	10,67	8,00	2,67 (33,38%)	390	390	0
2001	2	2	0	16	16	0	8,00	8,00	0,00	241	241	0
2000	2	2	0	19	19	0	9,50	9,50	0,00	175	175	0
1999	2	2	0	20	18	2 (11,11%)	10,00	9,00	1,00 (11,11%)	237	237	0
1998	2	2	0	16	16	0	8,00	8,00	0,00	160	160	0
1997	1	1	0	12	9	3 (33,33%)	12,00	9,00	3,00 (33,33%)	117	117	0
1996	2	2	0	17	16	1 (6,25%)	8,50	8,00	0,50 (6,25%)	205	205	0
1995	2	2	0	15	14	1 (7,14%)	7,50	7,00	0,50 (7,14%)	195	195	0

Fonte: consulta ao banco de dados⁸³

⁸³ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 81 – Dados comparativos da Coleção Saúde Pública: ISSN 0213-9111

0213-9111	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	69	70	-1 (-1,43%)	1.006	851	155 (18,21%)	14,58	12,16	2,42 (19,91%)	24.235	24.534	-299 (-1,22%)
2013	1	1	0	<u>25</u>	15	10 (66,67%)	25,00	15,00	10,00 (66,67%)	426	426	0
2012	6	6	0	<u>125</u>	85	40 (47,06%)	20,83	14,17	6,66 (47,00%)	2.751	2.751	0
2011	6	6	0	<u>114</u>	80	34 (42,50%)	19,00	13,33	5,67 (42,54%)	2.678	2.678	0
2010	6	6	0	<u>111</u>	78	33 (42,31%)	18,50	13,00	5,50 (42,31%)	2.383	2.383	0
2009	6	6	0	<u>132</u>	93	39 (41,94%)	22,00	15,50	6,50 (41,94%)	3.136	3.136	0
2008	6	6	0	<u>118</u>	79	39 (49,37%)	19,67	13,17	6,50 (49,35%)	2.339	2.339	0
2007	6	6	0	<u>111</u>	67	44 (65,67%)	18,50	11,17	7,33 (65,62%)	1.639	1.639	0
2006	6	6	0	<u>109</u>	63	46 (73,02%)	18,17	10,50	7,67 (73,05%)	1.521	1.521	0
2005	6	6	0	<u>103</u>	58	45 (77,59%)	17,17	9,67	7,50 (77,56%)	1.527	1.527	0
2004	7	8	-1 (-12,50%)	<u>132</u>	102	30 (29,41%)	18,86	12,75	6,11 (47,92%)	2.485	2.784	-299 (-10,74%)
2003	7	7	0	<u>120</u>	73	47 (64,38%)	17,14	10,43	6,71 (64,33%)	1.843	1.843	0
2002	6	6	0	<u>102</u>	58	44 (75,86%)	17,00	9,67	7,33 (75,80%)	1.507	1.507	0

Fonte: consulta ao banco de dados⁸⁴

⁸⁴ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 82 – Dados comparativos da Coleção Saúde Pública: ISSN 0124-0064

0124-0064	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	54	47	7 (14,89%)	643	575	68 (11,83%)	11,91	12,23	-0,32 (-2,62%)	15.726	13.949	1.777 (12,74%)
2012	4	4	0	<u>66</u>	50	16 (32,00%)	16,50	12,50	4,00 (32,00%)	1.546	1.228	318 (25,90%)
2011	6	6	0	<u>88</u>	87	1 (1,15%)	14,67	14,50	0,17 (1,17%)	2.082	2.099	-17 (-0,81%)
2010	7	8	-1 (-12,50%)	<u>122</u>	100	22 (22,00%)	17,43	12,50	4,93 (39,44%)	2.605	2.625	-20 (-0,76%)
2009	6	6	0	<u>95</u>	89	6 (6,74%)	15,83	14,83	1,00 (6,74%)	2.208	2.208	0
2008	6	6	0	<u>90</u>	88	2 (2,27%)	15,00	14,67	0,33 (2,25%)	2.190	2.190	0
2007	4	4	0	<u>60</u>	54	6 (11,11%)	15,00	13,50	1,50 (11,11%)	1.168	1.172	-4 (-0,34%)
2006	4	5	-1 (-20,00%)	<u>52</u>	52	0	13,00	10,40	2,60 (25,00%)	1.246	1.003	243 (24,23%)
2005	3	3	0	<u>28</u>	28	0	9,33	9,33	0,00	646	646	0
2004	4	4	0	<u>22</u>	22	0	5,50	5,50	0,00	660	660	0
2003	3	1	2 (200,00%)	<u>18</u>	5	13 (260,00%)	6,00	5,00	1,00 (20,00%)	482	118	364 (308,47%)
2002	4	0	4	<u>24</u>	0	24	6,00	0,00	6,00	427	0	427
2001	3	0	3	<u>17</u>	0	17	5,67	0,00	5,67	466	0	466

Fonte: consulta ao banco de dados⁸⁵

⁸⁵ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 83 – Dados comparativos da Coleção Saúde Pública: ISSN 0102-311X

0102-311x	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	199	219	-20 (-9,13%)	4.053	3.972	81 (2,04%)	20,37	18,14	2,23 (12,30%)	116.024	116.027	-3 (-0,00%)
2013	4	4	0	98	70	28 (40,00%)	24,50	17,50	7,00 (40,00%)	2.506	2.506	0
2012	13	13	0	273	228	45 (19,74%)	21,00	17,54	3,46 (19,73%)	7.423	7.423	0
2011	13	15	-2 (-13,33%)	313	264	49 (18,56%)	24,08	17,60	6,48 (36,82%)	8.752	8.752	0
2010	12	12	0	270	214	56 (26,17%)	22,50	17,83	4,67 (26,19%)	6.872	6.872	0
2009	13	15	-2 (-13,33%)	309	301	8 (2,66%)	23,77	20,07	3,70 (18,44%)	9.945	9.945	0
2008	13	16	-3 (-18,75%)	418	347	71 (20,46%)	32,15	21,69	10,46 (48,22%)	10.974	10.974	0
2007	13	16	-3 (-18,75%)	433	338	95 (28,11%)	33,31	21,12	12,19 (57,72%)	10.405	10.405	0
2006	13	13	0	343	260	83 (31,92%)	26,38	20,00	6,38 (31,90%)	7.845	7.845	0
2005	7	7	0	264	214	50 (23,36%)	37,71	30,57	7,14 (23,36%)	5.923	5.923	0
2004	7	8	-1 (-12,50%)	275	215	60 (27,91%)	39,29	26,88	12,41 (46,17%)	6.180	6.180	0
2003	7	8	-1 (-12,50%)	257	229	28 (12,23%)	36,71	28,62	8,09 (28,27%)	6.013	6.013	0
2002	7	7	0	216	189	27 (14,29%)	30,86	27,00	3,86 (14,30%)	5.728	5.728	0
2001	7	7	0	218	158	60 (37,97%)	31,14	22,57	8,57 (37,97%)	4.765	4.765	0
2000	5	6	-1 (-16,67%)	169	131	38 (29,01%)	33,80	21,83	11,97 (54,83%)	3.776	3.776	0
1999	5	6	-1 (-16,67%)	110	110	0	22,00	18,33	3,67 (20,02%)	3.083	3.083	0
1998	5	7	-2 (-28,57%)	187	131	56 (42,75%)	37,40	18,71	18,69 (99,89%)	3.082	3.082	0
1997	5	6	-1 (-16,67%)	138	96	42 (43,75%)	27,60	16,00	11,60 (72,50%)	2.312	2.312	0
1996	5	6	-1 (-16,67%)	103	77	26 (33,77%)	20,60	12,83	7,77 (60,56%)	1.778	1.778	0
1995	4	4	0	99	54	45 (83,33%)	24,75	13,50	11,25 (83,33%)	1.253	1.253	0
1994	5	6	-1 (-16,67%)	108	73	35 (47,95%)	21,60	12,17	9,43 (77,49%)	1.857	1.857	0
1993	5	5	0	92	56	36 (64,29%)	18,40	11,20	7,20 (64,29%)	1.336	1.336	0
1992	4	4	0	64	46	18 (39,13%)	16,00	11,50	4,50 (39,13%)	887	887	0
1991	4	4	0	48	38	10 (26,32%)	12,00	9,50	2,50 (26,32%)	869	869	0
1990	4	4	0	49	29	20 (68,97%)	12,25	7,25	5,00 (68,97%)	558	558	0
1989	4	4	0	36	18	18 (100,00%)	9,00	4,50	4,50 (100,00%)	269	269	0
1988	4	4	0	17	17	0	4,25	4,25	0,00	407	407	0
1987	4	4	0	33	24	9 (37,50%)	8,25	6,00	2,25 (37,50%)	445	445	0
1986	4	4	0	40	20	20 (100,00%)	10,00	5,00	5,00 (100,00%)	374	377	-3 (-0,80%)
1985	4	4	0	36	25	11 (44,00%)	9,00	6,25	2,75 (44,00%)	407	407	0

Fonte: consulta ao banco de dados⁸⁶

⁸⁶ Disponível em: <http://www.mcma.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>. Acesso em: 24 jun. 2013.

FIGURA 84 – Dados comparativos da Coleção Saúde Pública: ISSN 0042-9686

0042-9686	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	160	160	0	1.713	1.437	276 (19,21%)	10,71	8,98	1,73 (19,26%)	47.033	47.406	-373 (-0,79%)
2013	4	4	0	64	30	34 (113,33%)	16,00	7,50	8,50 (113,33%)	1.299	1.299	0
2012	12	12	0	198	95	103 (108,42%)	16,50	7,92	8,58 (108,33%)	3.537	3.537	0
2011	12	12	0	199	93	106 (113,98%)	16,58	7,75	8,83 (113,94%)	3.389	3.389	0
2010	12	12	0	236	103	133 (129,13%)	19,67	8,58	11,09 (129,25%)	3.742	3.751	-9 (-0,24%)
2009	12	12	0	241	105	136 (129,52%)	20,08	8,75	11,33 (129,49%)	3.520	3.549	-29 (-0,82%)
2008	12	12	0	245	105	140 (133,33%)	20,42	8,75	11,67 (133,37%)	3.820	3.930	-110 (-2,80%)
2007	12	12	0	212	107	105 (98,13%)	17,67	8,92	8,75 (98,09%)	3.030	3.127	-97 (-3,10%)
2006	12	12	0	232	111	121 (109,01%)	19,33	9,25	10,08 (108,97%)	2.981	3.018	-37 (-1,23%)
2005	12	12	0	211	102	109 (106,86%)	17,58	8,50	9,08 (106,82%)	3.281	3.363	-82 (-2,44%)
2004	12	12	0	204	105	99 (94,29%)	17,00	8,75	8,25 (94,29%)	3.160	3.160	0
2003	12	12	0	194	100	94 (94,00%)	16,17	8,33	7,84 (94,12%)	3.034	3.034	0
2002	12	12	0	240	121	119 (98,35%)	20,00	10,08	9,92 (98,41%)	3.583	3.583	0
2001	12	12	0	251	130	121 (93,08%)	20,92	10,83	10,09 (93,17%)	3.533	3.533	0
2000	12	12	0	245	130	115 (88,46%)	20,42	10,83	9,59 (88,55%)	5.124	5.133	-9 (-0,18%)

Fonte: consulta ao banco de dados⁸⁷

⁸⁷ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 85 – Dados comparativos da Coleção Saúde Pública: ISSN 0036-3634

0036-3634	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	80	127	-47 (-37,01%)	1.131	1.279	-148 (-11,57%)	14,14	10,07	4,07 (40,41%)	34.792	43.387	-8.595 (-19,81%)
2012	6	7	-1 (-14,29%)	<u>84</u>	70	14 (20,00%)	14,00	10,00	4,00 (40,00%)	2.217	2.529	-312 (-12,34%)
2011	7	10	-3 (-30,00%)	<u>154</u>	101	53 (52,48%)	22,00	10,10	11,90 (117,82%)	3.769	3.769	0
2010	7	8	-1 (-12,50%)	<u>127</u>	93	34 (36,56%)	18,14	11,62	6,52 (56,11%)	2.691	2.963	-272 (-9,18%)
2009	7	10	-3 (-30,00%)	<u>177</u>	119	58 (48,74%)	25,29	11,90	13,39 (112,52%)	5.088	5.088	0
2008	7	10	-3 (-30,00%)	<u>149</u>	104	45 (43,27%)	21,29	10,40	10,89 (104,71%)	3.966	3.963	3 (0,08%)
2007	7	10	-3 (-30,00%)	<u>126</u>	97	29 (29,90%)	18,00	9,70	8,30 (85,57%)	3.117	3.117	0
2006	7	8	-1 (-12,50%)	<u>105</u>	85	20 (23,53%)	15,00	10,62	4,38 (41,24%)	2.832	2.832	0
2005	6	6	0	<u>75</u>	46	29 (63,04%)	12,50	7,67	4,83 (62,97%)	1.472	1.472	0
2004	6	6	0	<u>84</u>	45	39 (86,67%)	14,00	7,50	6,50 (86,67%)	1.496	1.496	0
2003	7	11	-4 (-36,36%)	<u>165</u>	120	45 (37,50%)	23,57	10,91	12,66 (116,04%)	3.802	3.802	0
2002	7	7	0	<u>96</u>	78	18 (23,08%)	13,71	11,14	2,57 (23,07%)	2.409	2.409	0
2001	6	6	0	<u>81</u>	61	20 (32,79%)	13,50	10,17	3,33 (32,74%)	1.933	1.933	0

Fonte: consulta ao banco de dados⁸⁸

⁸⁸ Disponível em: <<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.

FIGURA 86 – Dados comparativos da Coleção Saúde Pública

0034-8910	FASCICULOS			ARTIGOS			MEDIA			CITAÇÕES		
	ANO	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO	DIFERENÇA	BC	SCIELO
TOTAL	253	261	-8 (-3,07%)	3.473	3.486	-13 (-0,37%)	13,73	13,36	0,37 (2,77%)	67.157	67.157	0
2012	7	7	0	153	138	15 (10,87%)	21,86	19,71	2,15 (10,91%)	3.043	3.043	0
2011	6	6	0	140	137	3 (2,19%)	23,33	22,83	0,50 (2,19%)	3.168	3.168	0
2010	6	6	0	132	132	0	22,00	22,00	0,00	2.764	2.764	0
2009	7	8	-1 (-12,50%)	177	157	20 (12,74%)	25,29	19,62	5,67 (28,90%)	3.056	3.056	0
2008	7	8	-1 (-12,50%)	187	170	17 (10,00%)	26,71	21,25	5,46 (25,69%)	3.257	3.257	0
2007	7	8	-1 (-12,50%)	188	157	31 (19,75%)	26,86	19,62	7,24 (36,90%)	3.048	3.048	0
2006	8	8	0	161	161	0	20,13	20,12	0,01 (0,05%)	3.285	3.285	0
2005	6	6	0	131	131	0	21,83	21,83	0,00	2.378	2.378	0
2004	7	7	0	129	129	0	18,43	18,43	0,00	2.254	2.254	0
2003	6	6	0	115	115	0	19,17	19,17	0,00	2.072	2.072	0
2002	6	7	-1 (-14,29%)	121	121	0	20,17	17,29	2,88 (16,66%)	2.360	2.360	0
2001	6	6	0	87	87	0	14,50	14,50	0,00	1.325	1.325	0
2000	6	7	-1 (-14,29%)	112	112	0	18,67	16,00	2,67 (16,69%)	1.874	1.874	0
1999	6	6	0	79	79	0	13,17	13,17	0,00	1.479	1.479	0
1998	6	6	0	85	85	0	14,17	14,17	0,00	1.883	1.883	0
1997	6	7	-1 (-14,29%)	96	88	8 (9,09%)	16,00	12,57	3,43 (27,29%)	2.080	2.080	0
1996	6	6	0	78	78	0	13,00	13,00	0,00	1.929	1.929	0
1995	6	6	0	69	69	0	11,50	11,50	0,00	1.573	1.573	0
1994	6	6	0	65	65	0	10,83	10,83	0,00	1.599	1.599	0
1993	7	7	0	65	65	0	9,29	9,29	0,00	1.592	1.592	0
1992	6	6	0	64	64	0	10,67	10,67	0,00	1.346	1.346	0
1991	6	6	0	68	68	0	11,33	11,33	0,00	1.578	1.578	0
1990	7	7	0	67	67	0	9,57	9,57	0,00	1.629	1.629	0
1989	6	6	0	62	62	0	10,33	10,33	0,00	1.187	1.187	0
1988	6	6	0	81	61	20 (32,79%)	13,50	10,17	3,33 (32,74%)	1.311	1.311	0
1987	6	6	0	66	51	15 (29,41%)	11,00	8,50	2,50 (29,41%)	1.100	1.100	0
1986	6	6	0	53	53	0	8,83	8,83	0,00	865	865	0
1985	6	6	0	56	56	0	9,33	9,33	0,00	976	976	0
1984	7	7	0	66	53	13 (24,53%)	9,43	7,57	1,86 (24,57%)	1.020	1.020	0
1983	6	6	0	40	40	0	6,67	6,67	0,00	1.084	1.084	0
1982	7	7	0	39	34	5 (14,71%)	5,57	4,86	0,71 (14,61%)	664	664	0
1981	7	7	0	70	64	6 (9,38%)	10,00	9,14	0,86 (9,41%)	1.239	1.239	0
1980	4	4	0	52	52	0	13,00	13,00	0,00	1.108	1.108	0
1979	5	6	-1 (-16,67%)	48	44	4 (9,09%)	9,60	7,33	2,27 (30,97%)	688	688	0
1978	4	4	0	51	51	0	12,75	12,75	0,00	836	836	0
1977	4	4	0	54	54	0	13,50	13,50	0,00	734	734	0
1976	5	6	-1 (-16,67%)	46	40	6 (15,00%)	9,20	6,67	2,53 (37,93%)	605	605	0
1975	4	4	0	60	55	5 (9,09%)	15,00	13,75	1,25 (9,09%)	777	777	0
1974	5	5	0	48	45	3 (6,67%)	9,60	9,00	0,60 (6,67%)	513	513	0
1973	4	4	0	43	41	2 (4,88%)	10,75	10,25	0,50 (4,88%)	435	435	0
1972	4	4	0	48	43	5 (11,63%)	12,00	10,75	1,25 (11,63%)	310	310	0
1971	2	2	0	32	28	4 (14,29%)	16,00	14,00	2,00 (14,29%)	294	294	0
1970	2	2	0	24	24	0	12,00	12,00	0,00	178	178	0
1969	2	2	0	28	26	2 (7,69%)	14,00	13,00	1,00 (7,69%)	199	199	0
1968	2	2	0	13	13	0	6,50	6,50	0,00	261	261	0
1967	2	2	0	24	21	3 (14,29%)	12,00	10,50	1,50 (14,29%)	201	201	0

Fonte: consulta ao banco de dados⁸⁹

⁸⁹ Disponível em:

<<http://www.cmca.srv.br/prototipo/mostraresumo.php?x=2&ss=1784231714&id=e62450dcb4be7a7cb4740a551f5a0933>>. Acesso em: 24 jun. 2013.