

ANDRÉS VILLAFUERTE OYOLA

**MODELAGEM PARA ORGANIZAÇÃO E REPRESENTAÇÃO
DO CONHECIMENTO EM ONTOLOGIAS DE DOMÍNIO:**

Uma experiência na área da cultura do sorgo

Belo Horizonte - MG
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

2010

ANDRÉS VILLAFUERTE OYOLA

**MODELAGEM PARA ORGANIZAÇÃO E REPRESENTAÇÃO
DO CONHECIMENTO EM ONTOLOGIAS DE DOMÍNIO:
Uma experiência na área da cultura do sorgo**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Ciência da Informação.

Área de concentração: Organização e Uso da Informação

Orientador: Profa. Dra. Lídia Alvarenga

Belo Horizonte
Escola de Ciência da Informação da UFMG
2010

Villafuerte, Andrés.

O98m Modelagem para organização e representação do conhecimento em ontologias de domínio [manuscrito]: uma experiência na área da cultura do sorgo / Andrés Villafuerte. – 2010.
180 f. : il., enc.

Orientadora: Lídia Alvarenga.

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

Referências: f. 123-129

Apêndices: f. 130-167

Anexos: f. 168-180

1. Ciência da Informação – Teses. 2. Representação do conhecimento (Teoria da informação) – Teses. 3. Ontologias (Recuperação da informação) – Teses. 4. Pesquisa Agropecuária – Teses. 5. Sorghum bicolor – Teses. I. Título. II. Alvarenga, Lídia. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 025.4.03

DEDICATÓRIA

Ao Deus eterno, pai de toda a ciência. Pela fé e ânimo em todos os momentos difíceis desta travessia.

A Mônica, minha esposa, e aos nossos filhos Roger e Caroline, que em muitos momentos, e realmente foram muitos mesmo, se viram privados da minha presença.
Espero agora poder recompensá-los.

Ao meu inesquecível amigo canino *Spyke*, que entrou na minha vida no início do doutorado, e durante as muitas horas de solidão que a pesquisa nos impõe, foi meu companheiro fiel e solidário. Havendo cumprido sua missão, ele partiu no fim deste período de doutorado, deixando saudades imensas!

AGRADECIMENTOS

À professora Lídia Alvarenga, cujo papel nestes últimos anos foi grande demais, de modo a não caber mais na palavra orientadora. Agradeço pelos conhecimentos, dedicação e paciência, transformando alguém que tendo começado com grande interesse, terminou o doutorado como amante incondicional da Ciência da Informação.

Ao Programa de Pós-Graduação em Ciência da Informação da UFMG, que através da arte de fazer ciência, me ajudaram a criar uma nova consciência acadêmica e científica, que hoje busco com afinco aplicar na docência.

Ao professor Ricardo Barbosa, pela elegância das suas exposições teóricas, que me despertaram um profundo interesse pelo conhecimento organizacional.

Aos professores Renato Souza e Maurício Almeida que possuem um perfil singular, associando com o mesmo grau de intensidade, competência profissional e humildade pessoal.

À direção do Centro Universitário de Sete Lagoas, que em momento oportuno me concedeu licença, viabilizando a conclusão desta pesquisa.

Aos pesquisadores da Embrapa, engenheiros agrônomos Flávio Dessaune Tardin e José Avelino Rodrigues, que a despeito das suas demandas profissionais, mostraram disponibilidade quando precisei e me introduziram no rico universo da cultura do sorgo.

A Victor Miguel Paredes, doutorando da UFMG, pela sua hospitalidade sem limites. Nas minhas inúmeras idas e vindas de Sete Lagoas até a UFMG, me hospedou na sua casa, cuja localização privilegiada próxima ao campus, foi estratégica para este pesquisador itinerante. Victor, você não pode imaginar como esse gesto foi importante.

A todas as pessoas, que seria impossível mencionar sem pecar por omissão, que durante estes últimos anos, de forma direta ou indireta, e muitas vezes sem saber, foram importantes dentro do meu contexto pessoal.

À CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pelo auxílio financeiro concedido durante o período de desenvolvimento desta pesquisa. A ela manifesto o meu sincero desejo de que dias melhores virão, e com eles, um maior volume de investimentos em pesquisas no Brasil.

"Deus deliberadamente escolheu valer-se de idéias que o mundo considera absurdas e desprezíveis para envergonhar aqueles indivíduos que o mundo considera sábios e grandes"

(1 Coríntios, 1:27)

"Criamos a época da velocidade, mas nos sentimos enclausurados dentro dela. Nossos conhecimentos fizeram-nos céticos; nossa inteligência, empedernidos e cruéis. Pensamos em demasia e sentimos bem pouco. Mais do que máquinas, precisamos de humanidade"

Charles Chaplin (1889 – 1977)

"A percepção do desconhecido é a mais fascinante das experiências. O homem que não tem os olhos abertos para o misterioso passará pela vida sem ver nada"

Albert Einstein (1879 - 1955)

SUMÁRIO

1. INTRODUÇÃO.....	1
1.1 Contexto social, tecnológico e econômico no qual se desenvolve a pesquisa.....	3
1.2 Contexto da pesquisa no âmbito da Ciência da Informação.....	5
1.3 Inserção da pesquisa no âmbito da representação do conhecimento e do estudo e desenvolvimento de ontologias	6
1.4 Delimitação do problema de pesquisa.....	9
1.5 Análise de ontologias desenvolvidas no setor de agronomia.....	11
1.5.1 A Árvore do Conhecimento.....	12
1.6 Definição dos objetivos de pesquisa.....	17
2. FUNDAMENTAÇÃO TEÓRICA, CONCEITUAL E CONTEXTUAL DA PESQUISA.....	19
2.1 O conhecimento.....	21
2.1.1 O conhecimento científico.....	26
2.1.1.1 Critérios de avaliação de pesquisas em comunidades científicas.....	28
2.1.2 A representação do conhecimento.....	30
2.2 Organização e representação do conhecimento através de estruturas semânticas	36
2.2.1 Método Analítico-Sintético.....	36
2.2.2 Teoria do Conceito.....	37
2.2.3 Análise de Assunto.....	39
2.3 Fundamentos teóricos da classificação.....	40
2.3.1 Categorização.....	42
2.3.2 Cânones para o trabalho no plano das idéias.....	44
2.4 Compromisso Ontológico e Epistemológico.....	46
2.5 Ontologias.....	49
2.5.1 Revisitando o conceito de ontologias.....	51
2.5.2 Finalidades e aplicabilidades das ontologias.....	57
2.5.3 Tipologias.....	58
2.5.4 Metodologias, métodos e técnicas para o desenvolvimento de ontologias.....	60
2.6. A pesquisa científica no Brasil.....	69
2.6.1 Institutos de pesquisas agrônomicas no Brasil.....	71
2.6.1.1 Embrapa Milho e Sorgo.....	73
2.7 Implicações para a pesquisa.....	74
3. METODOLOGIA DE PESQUISA.....	76
3.1 Contexto institucional da pesquisa.....	76
3.2 Descrevendo a natureza e tipologia da pesquisa.....	78
3.3 Considerações sobre modelagem.....	81
3.4 - Etapas da abordagem proposta para modelagem do conhecimento científico.....	84

3.4.1 – Levantamento de insumos.....	84
3.4.1.1 Pesquisadores.....	85
3.4.1.2 Fontes.....	87
3.4.1.3 Veículos utilizados na comunicação científica.....	87
3.4.1.4 Produção literária.....	90
3.4.2 Análise dos artigos científicos.....	93
3.4.2.1 Divisão do texto em partes menores.....	94
3.4.2.2 Divisão do texto identificando enunciados significativos.....	97
3.4.3 Categorização do domínio utilizando o método analítico sintético.....	98
3.4.4 Avaliação final do modelo conceitual junto aos especialistas.....	100
4. RESULTADOS: ANÁLISE E DISCUSSÃO.....	102
5. CONSIDERAÇÕES FINAIS E ESTUDOS FUTUROS.....	118
REFERÊNCIAS	123
APÊNDICE 1 - Estrutura Semântica Representativa do Conhecimento Científico da Cultura do Sorgo.....	130
APÊNDICE 2 - Trechos extraídos de cada um dos artigos utilizados na elaboração da modelagem.....	148
APÊNDICE 3 - Glossário de termos e conceitos do segmento agropecuário.....	164
ANEXO 1 - Artigos utilizados na elaboração da modelagem.....	168
ANEXO 2 - Dissertações e Teses desenvolvidas no Brasil, no escopo da Ciência da Informação, envolvendo o assunto ontologias.....	178
ANEXO 3 - Institutos de Pesquisas Agronômicas Brasileiros.....	180

LISTA DE FIGURAS

Figura 1 - Árvore do Conhecimento do trigo.....	14
Figura 2 - Detalhes dos nós principais da Árvore.....	15
Figura 3 - Representação de uma rede semântica.....	34
Figura 4 - Principais passos do método de Uschold e King.....	62
Figura 5 - Principais passos da metodologia de Grüninger e Fox.....	63
Figura 6 - Ciclo de vida de uma ontologia na METHONTOLOGY.....	67
Figura 7 - Planta do Sorgo.....	102
Figura 8 - Mapa geral do conhecimento científico do sorgo.....	103
Figura 9 - Mapa conceitual da classe Sorgo.....	106
Figura 10 - Mapa conceitual da classe Genética e Melhoramento.....	107
Figura 11 - Mapa conceitual da classe Manejo da Cultura.....	109
Figura 12 – Mapa conceitual da classe Alimentação.....	114
Figura 13 – Mapa conceitual da classe Pesquisa Agropecuária.....	115
Figura 14 – Mapa conceitual da classe Pessoas.....	116

LISTA DE TABELAS

Tabela 1 - Tipos de Ontologias.....	59
Tabela 2 - Pesquisadores do sorgo na Embrapa.....	86
Tabela 3 - Comunicação científica via Periódicos.....	89
Tabela 4 - Comunicação científica via Congressos.....	89
Tabela 5 - Totais de trechos selecionados por artigo.....	96

LISTA DE ABREVIATURAS E SIGLAS

ASRG	Área de superfície de raiz muito fina
AST	Área de superfície total
CRF	Comprimento da Raiz Final
CRG	Comprimento de Raiz Grossa
CRI	Comprimento da Raiz Inicial
CRMF	Comprimento de raiz muito fina
CRRS	Comprimento Relativo de Raiz Seminal
CTR	Comprimento total de raiz
DAMS	digestibilidade aparente da matéria seca
DAPB	digestibilidade aparente da proteína bruta
DIVMS	digestibilidade in vitro da matéria seca
DMR	Diâmetro médio de raiz
DTR	Densidade de tecido de raiz
EDA	energia digestível aparente
EMA	energia metabolizável aparente
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
EMV	energia metabolizável verdadeira
FDA	fibra em detergente ácido
FDN	fibra em detergente neutro
FR	Finura de raiz
MS	matéria seca
MSD	matéria seca digestível
MSMA	matéria seca metabolizável aparente
MV	matéria verde
N-NH ₃ /NT	nitrogênio amoniacal / nitrogênio total
PFPA	peso fresco da parte aérea
PFR	peso fresco da raiz
pH	abreviação de "potencial hidrogeniônico", que é uma escala usada para medir a acidez ou a alcalinidade de soluções evitando o uso de expoentes, através da medida de concentração do íon hidrogênio em solução. O pH abaixo de 7 é ácido; acima de 7, alcalino; e 7 é considerado neutro.
PSPA	peso seco de parte aérea da planta
PSR	peso seco da raiz da planta
PSRA	peso seco de raiz
PSRA/PSPA	relação peso seco raiz/peso seco parte aérea
VTR	Volume total de raiz

VILLAFUERTE, Andrés. *Modelagem para organização e representação do conhecimento em ontologias de domínio: uma experiência na área da cultura do sorgo*. Tese de Doutorado - Universidade Federal de Minas Gerais, Escola de Ciência da Informação, 2010.

RESUMO

Propõe-se uma modelagem para a organização e representação do conhecimento de um dado domínio, utilizando o ferramental da Ciência da Informação. O objetivo principal desta pesquisa de doutorado é apresentar uma contribuição para o campo das ontologias de domínio, através de uma abordagem que possibilite implementar a etapa de mapeamento do conhecimento de um domínio específico. Após analisar diversos métodos de construção de ontologias de domínio, amplamente divulgados na literatura técnica, foi verificado que a fase de mapeamento do conhecimento a qual procura levantar e representar o conhecimento do especialista é descrita em geral com poucos detalhes ou de forma muito sucinta, dificultando a compreensão e aplicação da técnica. A utilização de técnicas oriundas da Ciência da Informação, tais como: teoria do conceito, método analítico-sintético, teoria da classificação, somadas aos princípios da garantia literária, cultural e científica e às diretrizes que norteiam o compromisso ontológico, apresentaram em conjunto resultados satisfatórios, através de uma modelagem, que ao ser apresentada em detalhes, viabiliza sua aplicação no mapeamento do conhecimento de qualquer domínio. O universo empírico de dados compreendeu 18 artigos científicos da frente de pesquisa da cultura do sorgo dos quais foram extraídos 351 expressões significativas que deram origem a 672 termos que denotam conceitos da área. Como forma de testar e validar a modelagem proposta, a mesma foi utilizada para mapear o conhecimento científico gerado a partir das pesquisas com sorgo, cereal utilizado na alimentação de animais, desenvolvidas pela Embrapa - Empresa Brasileira de Pesquisa Agropecuária, na sua unidade de pesquisa localizada em Sete Lagoas – MG. O produto desta tese é apresentado na forma de uma estrutura semântica que reflete o universo de pesquisas e do conhecimento científico da cultura do sorgo.

Palavras-chave: Representação do Conhecimento. Método Analítico-Sintético. Estrutura Semântica. Ontologias. *Sorghum bicolor*. Pesquisa Agropecuária.

VILLAFUERTE, Andrés. *Modelado para la organización y representación del conocimiento em ontologías de dominio: experiencia en el área de sorgo*. Tesis de Doctorado - Universidade Federal de Minas Gerais, Escola de Ciencia de la Información, 2010.

RESUMEN

Se propone un modelado para organización y representación del conocimiento de cualquier dominio, utilizando las herramientas de la Ciencia de la Información. Esta investigación de doctorado tiene como objetivo principal ofrecer una contribución al campo de las ontologías de dominio, a través de un enfoque que permite desarrollar el mapeo del conocimiento de un dominio específico. Después de analizar varios métodos de construcción de ontologías de dominio, ampliamente descritos en la literatura técnica, se comprobó que la etapa de mapeo del conocimiento, que busca crear y representar el conocimiento de expertos, se suele describir con pocos detalles o de manera muy sucinta. Debido a esto se hace difícil el entendimiento y la aplicación de la técnica. La utilización de técnicas de la Ciencia de la Información, tales como: Teoría del Concepto, Método Analítico-Sintético, Teoría de la Clasificación, además de los principios de garantía literaria, científica y cultural y de las directrices para obtener el compromiso ontológico, mostrarón en conjunto, resultados satisfactorios. Este modelo que se presenta en detalle permite su aplicación para el mapeo del conocimiento de cualquier dominio. El universo empírico de los datos incluyó 18 artículos científicos de investigación avanzada sobre el sorgo, de los cuales fueron extraídos 351 expresiones significativas, lo que resultó en 672 términos, que denotan conceptos en este ámbito. Para probar y validar el modelo propuesto, se utiliza este modelo para mapear los conocimientos científicos generados por la investigación de sorgo, cereal utilizado para la alimentación animal, realizada por la Embrapa - Empresa Brasileira de Pesquisa Agropecuária, en su unidad de investigación situada en Sete Lagoas – Estado de Minas Gerais – Brasil. El producto de esta tesis es presentado bajo la forma de una estructura semántica que representa el universo de la investigación y el conocimiento científico de sorgo.

Palabras clave: Representación del Conocimiento. Método Analítico-Sintético. Estructura Semántica. Ontología. Sorgo. Investigación Agronómica.

VILLAFUERTE, Andrés. *Modeling for organization and knowledge representation in domain ontologies: experience in the sorghum area*. Doctoral Thesis - Universidade Federal de Minas Gerais, School of Information Science, 2010.

ABSTRACT

This work proposes a model for organization and knowledge representation of any domain, using the tools of Information Science. This doctoral research has as its main objective to offer a contribution to the field of domain ontologies, through an approach that enables implementing the step of mapping knowledge in a specific domain. After analysing several methods of construction of domain ontologies, widely reported in the technical literature, was verified that the mapping stage of knowledge, which seeks to raise and represent the expert knowledge, is usually described in little detail or very succinctly. This fact makes difficult to understand and apply the technique. The use of techniques from Information Science, such as: Theory of Concept, Analytic-Synthetic Method, Theory of Classification, in addition to the principles of literary, cultural and scientific warranty and the guidelines for the Ontological Commitment, showed satisfactory results, all together. This modeling that is presented in details enables its application for the mapping of knowledge from any domain. The empirical universe of data included 18 scientific articles in the research front of *sorghum*, of which, 351 significant expressions were extracted, resulting into 672 terms, denoting concepts in this area. To test and validate the proposed model, it was used to map the scientific knowledge generated from *sorghum* research, cereal used in animal feeding, undertaken by the Embrapa - Empresa Brasileira de Pesquisa Agropecuária, in its research unit in Sete Lagoas city – State of Minas Gerais - Brazil. The product of this thesis is presented as a semantic structure representing the universe of research and scientific knowledge of sorghum.

Keywords: Knowledge Representation. Analytic-Synthetic Method. Semantic Structure. Ontologies. *Sorghum bicolor*. Agronomics Research.