

MAPEAMENTO E REPRESENTAÇÃO DO CONHECIMENTO E MODELAGEM ONTOLÓGICA: Uma experiência na área da cultura do sorgo

Andrés Villafuerte Oyola*
Lídia Alvarenga**

RESUMO

O presente trabalho destaca a importância do estudo das ontologias para a Ciência da Informação e percorre através de uma análise da literatura técnica, diversos métodos desenvolvidos para construção de ontologias. A partir desta análise e considerando que ontologias de domínio têm como objetivo principal mapear o conhecimento de uma área específica, verificou-se a partir da literatura estudada que esta fase de mapeamento onde procura-se levantar o conhecimento do especialista, apresenta em geral poucos detalhes ou é descrita de forma muito sucinta, dificultando a sua compreensão e utilização. Baseada nesta constatação, a pesquisa de doutorado aqui relatada, em andamento, assumiu a tarefa de preencher esta lacuna, desenvolvendo um método que descreve de forma clara e abrangente esta etapa de aquisição de conhecimento, onde através do levantamento de insumos importantes, busca-se estruturar o conhecimento de um domínio, que será implementado posteriormente em uma ontologia de domínio. O método em questão vem sendo utilizado para mapear o conhecimento científico gerado a partir das pesquisas de ponta desenvolvidas no setor agropecuário pela EMBRAPA Milho e Sorgo, localizada na cidade de Sete Lagoas - Minas Gerais.

Palavras-chave: Ontologia. Construção de Ontologias. Aquisição de Conhecimento. Pesquisa Agropecuária.

1 INTRODUÇÃO

Ontologias vêm conquistando um espaço cada vez maior no corpo de pesquisas da Ciência da Informação no Brasil. Esta tendência é justificada pela capacidade que as ontologias oferecem para a organização da informação e a possibilidade de estruturação do conhecimento de um dado domínio, objetivos relevantes para o referido campo de conhecimento.

Uma análise das pesquisas da área em nível de pós-graduação revela que a partir de 2003 (ver Anexo 1) houve um número expressivo de dissertações e teses que adotaram ontologias como seu objeto de pesquisa. O tema foi abordado sob suas diversas óticas, quais

* Doutorando em Ciência da Informação, Escola de Ciência da Informação - UFMG; Professor do Centro Universitário de Sete Lagoas, andres@eci.ufmg.br

** Professora Doutora Titular da Escola de Ciência da Informação - UFMG, lidiaalvarenga@eci.ufmg.br

sejam: estudos conceituais, avaliação de metodologias, modelos para construção de ontologias, análise de relações em ontologias, aplicações de *web* semântica, etc.

O presente artigo faz parte da pesquisa de doutorado atualmente em desenvolvimento no Programa de Pós-Graduação da ECI da UFMG, e retoma o assunto ontologias, buscando dar uma contribuição significativa nos aspectos referentes à etapa de mapeamento do conhecimento.

Ao analisar os diversos métodos para construção de ontologias, entre eles o projeto Cyc (LENAT, 1994), Enterprise Ontology (USCHOLD e GRUNINGER, 1996), 101 (NOY e GUINNESS, 2001) e as metodologias propostas pelo projeto TOVE (FOX e GRUNINGER, 1998), e a Methontology (FERNÁNDEZ-LÓPEZ et al, 1997), percebe-se que embora seja freqüente mencionarem a etapa referente ao mapeamento do domínio, que implica no processo de aquisição de conhecimento, não se encontram explicitados detalhes sobre esta etapa que permitam ter uma completa compreensão das técnicas utilizadas. A descrição desta etapa, invariavelmente é feita de forma muito vaga, o que não oferece condições de ter uma visão clara dos procedimentos adotados. Este trabalho de pesquisa apresenta um método que descreve de forma clara, ampla e detalhada algumas técnicas utilizadas para adquirir o conhecimento do domínio,

A abordagem aqui proposta parte de reflexões filosóficas sobre a natureza do conhecimento, e está ancorada em conceitos referentes a ontologia, modelagem ontológica e Compromisso Ontológico (SMITH, 2004). Utiliza ainda ferramentas como Garantía Literária (BARITÉ, 2007), Análise de Assunto (DIAS e NAVES, 2007; FUJITA, 2003) e princípios teóricos da classificação.

2 FUNDAMENTAÇÃO TEÓRICA

O estudo de ontologias ganha clareza se partirmos da análise da gênese desta palavra, e para isto é ponto de partida obrigatório abordar os princípios filosóficos sobre o qual ele repousa. Com este objetivo, iniciamos uma viagem ao passado que se inicia no ano 50 a.c, época em que segundo Chauí (2000, p. 266), Andrônico de Rodes dedicou esforços para a coleta e classificação da obra de Aristóteles (384-322 a.c.), que tinha se dispersado ao longo dos séculos. Ao fim deste trabalho o classificador decidiu que as obras de Aristóteles deveriam ficar localizadas após os tratados da física, razão pela qual cunhou o termo *Ta meta ta physika* ou simplesmente metafísica, para se referir aos escritos catalogados em uma ordem tal que os colocassem após os escritos da física, ou como estabelece o vocábulo grego, *Ta*: aqueles; *meta*: após, depois; *ta physika*: àqueles da física.

Porém, o próprio Aristóteles definiu que o assunto dos seus escritos deveria ser designado como Filosofia Primeira, cujo tema central é focado no “estudo do ser enquanto ser”, tendo prevalecido no entanto, o termo metafísica. Vale ressaltar aqui que a metafísica surgiu da necessidade dos filósofos estabelecerem um conhecimento do universo que fosse além daquele estabelecido pela cosmologia e pela física, e é dentro deste contexto que ela passa a ocupar um lugar central na filosofia.

O termo ontologia foi utilizado no século XVII pelo filósofo alemão Jacobus Thomasius, que utilizou o termo cunhado pelo filósofo escolástico alemão Rudolf Goclenius em 1613, estabelecendo que os estudos da metafísica ou Filosofia Primeira, como foram chamados por Aristóteles, fossem conhecidos sob este nome. Segundo Chauí (2000, p. 266), surge assim a palavra ontologia, que tem na sua composição a palavra *Onto*, derivada de dois substantivos gregos que na sua essência trazem o significado da busca do entendimento do Ser, tal como ele é e não como parece ser. Desta forma a palavra ontologia adquire significado como o estudo do Ser, das coisas ou dos entes, tais como são em si mesmas, real e verdadeiramente.

O estudo aqui apresentado ao tratar de ontologias segue esta orientação, ou seja, a nossa abordagem é focada no estudo do Ser das coisas e baseado nisto, pode-se afirmar que “uma ontologia representa a modelagem de um Ser específico, que ao ser analisado e estudado nos leva a compreender a sua essência, permitindo relacionar de maneira inequívoca, precisa e organizada as suas propriedades, características, e o seu comportamento”.

Percebe-se claramente que ao se fazer uma análise ontológica busca-se adquirir conhecimento sobre um Ser, e aqui poderia se fazer uma série de considerações filosóficas a respeito do conhecimento a ser obtido, tais como: Quem é o Ser? Qual a sua essência? Com que olhar devo me aproximar do Ser? Como deve se fazer a delimitação do Ser? ou Quais partes do Ser representam melhor o todo? Estas considerações foram analisadas e discutidas na pesquisa de doutorado, na qual se baseou este artigo mas não fazem parte do escopo do mesmo.

Um fato que deve ser entendido é que a construção do conhecimento humano passa invariavelmente pela categorização de conceitos e é por causa deste “*modus operandis*” do nosso cérebro que é comum quando nos deparamos com o novo, fazemos uma tentativa de classificar este elemento dentro de uma estrutura de conceitos pré-existente; mas caso isto não seja factível é criada uma nova categoria onde este elemento possa ser alocado. Portanto, podemos afirmar que o nosso fazer cognitivo, está ligado fortemente a estruturas taxonômicas

que, por sua vez são armazenadas nas nossas regiões da memória. A seguinte frase reforça o que foi afirmado acima:

Podemos ver que as categorias têm uma capacidade de estrutura: não apenas estruturam, de fato, todos os nossos elementos de conhecimento e unidades do conhecimento; elas fornecem, ao mesmo tempo, por este meio, o esqueleto, os ossos e tendões para estruturar todo o nosso conhecimento. (DAHLBERG, 1978, p.34).

Por outro lado, sabe-se que para que o conhecimento possa ser compartilhado entre humanos fazemos uso de palavras e símbolos e outros elementos característicos da linguagem, e isto vale tanto no caso do conhecimento mais simples como no conhecimento mais elaborado como aquele ostentado por especialistas de uma determinada área científica. Alinhando-se a esta lógica, o processo de construção de ontologias passa também necessariamente pela criação de um conjunto de conceitos e respectivos termos, palavras ou símbolos que representam o vocabulário utilizado para a comunicação entre os especialistas da área.

No entanto, a ação de especificar com maior ou menor profundidade os termos de um vocabulário, a partir de seus respectivos conceitos, conduziria no máximo à obtenção de um tesouro, estrutura que por si só não contempla as múltiplas possibilidades proporcionadas por uma ontologia. Portanto, para poder se chegar a uma modelagem ontológica o passo seguinte seria fazer o mapeamento entre termos e as coisas ou idéias que elas representam, e isto é conseguido utilizando-se relações e conceitos. Fazendo-se uma analogia com as estruturas cognitivas do ser humano, percebe-se que ontologias buscam emular o mapeamento que é realizado por meio das sinapses ou links que são realizadas continuamente pelos nossos neurônios, considerando que este mapeamento se dá através de conexões complexas que são criadas no cérebro humano.

É importante ressaltar que o momento em que acontecem estas conexões corresponde a um momento de intensa atividade intelectual representada por tarefas de alto nível de complexidade, tarefas que demandam raciocínio, dedução, lógica e inferências as quais durante muitos anos têm desafiado pesquisadores ligados à Neurociência, Ciência da Computação, Psicologia, Linguística que têm buscado compreender o comportamento do cérebro humano. A este grupo soma-se mais recentemente pesquisadores da Ciência da Informação que têm entre suas preocupações e interesses de pesquisa, os fenômenos cognitivos envolvidos no processamento da informação e construção do conhecimento.

A figura 1 resume esquematicamente as etapas acima citadas, que fazem parte do processo de estruturação de uma ontologia e que refletem a atividade cognitiva humana.

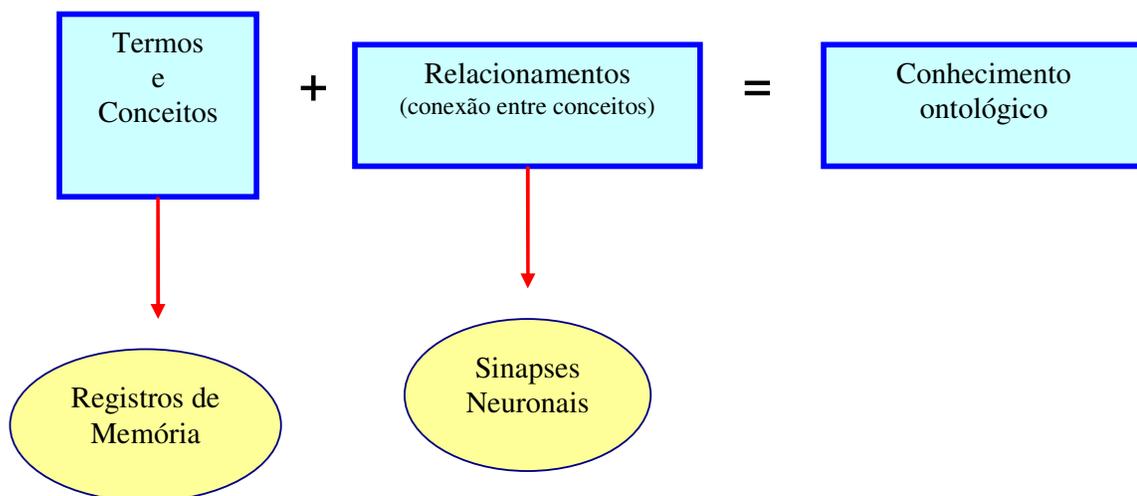


Figura 1 – Esquema geral de uma estrutura ontológica

A figura acima, embora defina esquematicamente uma estrutura ontológica, não deixa evidente o processo de mapeamento do conhecimento, ou seja, os passos que devem orientar o levantamento dos termos, a definição dos conceitos e a forma de se fazer estes relacionamentos, estes passos serão apresentados nas seções 3.1 a 3.3. Nas seções 2.1 a 2.3 serão descritos alguns suportes teóricos nos quais a presente pesquisa se apoiou, entre eles destacam-se os conceitos de compromisso ontológico, garantia literária e análise de assunto.

2.1 Compromisso ontológico

A noção de compromisso ontológico foi introduzida pelo filósofo americano Willard Quine numa série de ensaios importantes entre os quais figura o clássico “On What There Is” (QUINE, 1948). A noção de compromisso ontológico se baseia em dois aspectos:

a) validação de uma teoria que retrata uma realidade ou experiência e b) comprovação da existência dos objetos que fazem parte da ontologia da teoria.

Segundo proposto por Quine, uma teoria acerca de um determinado segmento da realidade ou da experiência é simplesmente uma coleção consistente de crenças ou afirmações acerca deste segmento em questão, expressas numa determinada linguagem. Esta teoria será verdadeira se todas as crenças que a compõem, logo se todas as conseqüências lógicas dessas crenças forem de fato verdadeiras.

Por outro lado, os objetos com os quais uma teoria está ontologicamente comprometida são precisamente aqueles objetos cuja existência é reconhecida e assumida pela teoria, seja de forma explícita ou implícita, tais objetos formam a ontologia da teoria, que se

constitui de um conjunto de entidades cuja inexistência traria como consequência imediata, a falsidade da teoria. Neste sentido Quine afirma que para que uma entidade seja assumida como tal, basta que ela seja considerada um valor ou ocorrência de uma variável. Por exemplo: afirmar que existem comerciantes desonestos não gera um comprometimento sobre a existência de comércio e desonestidade, no entanto, para tornar a declaração verdadeira basta apresentar um comerciante desonesto, representando assim o valor de uma variável.

Uma das propostas mais célebres de Quine consiste num processo para determinar com que objetos, ou com que classes ou categorias de objetos, uma dada teoria está ontologicamente comprometida. Este processo não permite determinar de uma forma ampla e irrestrita, o que há, ou o que existe. O processo por estar restrito ao espectro da teoria apenas permite verificar o que há, ou o que existe, para aquela teoria específica. Por estes motivos, torna-se relevante e significativo determinar com que objetos e com quais categorias de objetos está ontologicamente comprometido o nosso sistema de crenças.

A título de exemplificar o que foi proposto por Quine, pode-se imaginar o seguinte diálogo: uma pessoa ao mostrar para outra duas moedas, afirma que uma moeda possui brilho mais intenso do que a outra, a segunda pessoa retruca afirmando que o máximo que se pode afirmar é que as percepções visuais de ambos são diferentes, reivindicando-se que a análise correta deveria focar aspectos que envolvessem questões sensoriais referentes a sensações de claridade e ofuscamento. Desse diálogo imaginário, conclui-se que a primeira pessoa está comprometida com uma ontologia que envolve o objeto “moeda” e a entidade “brilho”, enquanto a segunda pessoa está comprometida com uma ontologia que envolve as diferenças de percepção visual. Pode-se ainda acrescentar o fato de que a entidade brilho, cuja existência é enfaticamente confirmada pela primeira pessoa e categoricamente negada pela outra, traz ainda o inconveniente de lidar com o termo brilhante, o que implica em estabelecer uma comparação entre duas entidades que são inexistentes para a segunda pessoa, tornando portanto sem validade a afirmação feita pela primeira. Elaborar declarações com variáveis ligadas não determina o que há, mas unicamente o que eu estou disposto a dizer que há; desde que os debates ontológicos tomam lugar em níveis lingüísticos e semânticos, a identificação dos comprometimentos ontológicos é crucial para o entendimento dos esquemas conceituais subjacentes ao debate.

Na pesquisa de doutorado o conceito de compromisso ontológico é utilizado como um parâmetro de referência, útil na validação do domínio mapeado, no caso o domínio referente à cultura do sorgo. A análise deste domínio dará origem a um conjunto de afirmações que

devem ser validadas e que em conjunto formam a teoria referente ao segmento agropecuário, em particular ao espectro de pesquisas do sorgo.

2.2 Garantia literária, cultural e científica

A garantia literária é um importante instrumento que dá suporte à construção de vocabulários controlados e segundo Wyndham Hulme, que introduziu o conceito em 1911, as classes utilizadas nas linguagens de indexação não deveriam ser definidas seguindo critérios de classificação do conhecimento e sim pelas classes encontradas na literatura, ou seja, a própria literatura deve servir como fonte para a especificação do vocabulário utilizado para representar os assuntos que a descreve.

A respeito do conceito de garantia literária, há também importantes contribuições apresentadas por Barité (2007), que fazendo uma revisão do conceito levanta questões relevantes tais como:

- Em quais documentos específicos deve se basear a garantia literária?
- Que critérios devem ser aplicados na análise de documentos e que vão permitir uma correta utilização do critério de garantia literária?
- É válido basear a garantia literária apenas na literatura mais atual ou deve-se considerar a literatura produzida ao longo do tempo?
- Que tipo de conhecimento deve estar legitimado pela garantia literária?

Algumas respostas às questões acima têm sido apresentadas por pesquisadores de renome, no âmbito da Ciência da Informação, tais como Vickery (1997), Gilchrist (2003) e o próprio Barité (2007). No decorrer da presente pesquisa, buscar-se-á fundamentar alguns critérios aqui adotados baseados nas propostas destes autores.

Por outro lado, Barité (2007) afirma que o princípio que sustenta a garantia literária, deixou há muito tempo de se concentrar nas análises unicamente do vocabulário da literatura a ser classificada, como originalmente proposto por Hulm, e acrescenta que, atualmente, o princípio foi ampliado de forma a incluir conceitos como garantia do usuário, garantia cultural e garantia organizacional.

2.3 Análise de Assunto

Segundo Dias e Naves (2007), a análise de assunto é considerada um dos processos básicos da organização da informação e consiste na compreensão e interpretação do conteúdo de um documento. É uma etapa que se desenvolve dentro das atividades de indexação, portanto, é um processo importante que impacta a eficiência da recuperação da informação.

Fujita (2003) afirma que a análise de assunto é composta de três estágios que se superpõem: a) compreensão do conteúdo do documento; b) identificação de conceitos relevantes que representam o documento e c) seleção de conceitos válidos que serão úteis na recuperação.

No primeiro estágio recomenda-se a leitura integral do documento, embora em alguns casos (livros, teses, monografias) isto seja impraticável, mesmo sendo a situação ideal. No caso da presente pesquisa conforme relatado na seção 3.3 foi realizada a leitura na íntegra dos 23 artigos científicos. O documento Sistema Internacional de Informação para Ciência e Tecnologia, elaborado pela UNESCO fornece algumas diretrizes em relação à prática da análise de assunto, recomendando atenção especial durante a leitura para o título, introdução, as primeiras frases de capítulos e parágrafos; ilustrações, tabelas, diagrama e suas explicações; conclusão; palavras ou grupos de palavras sublinhadas ou impressas com tipo diferente. Por outro lado destaca que os primeiros itens do texto apresentam, geralmente, as intenções do autor enquanto que as partes finais comunicam o alcance dessas intenções. Estas orientações foram seguidas durante a etapa de análise dos artigos científicos citados na seção 3.3.

No segundo estágio, o de identificação de conceitos, que demanda intensa atividade cognitiva, busca-se selecionar conceitos que melhor representem o conteúdo do texto, e neste momento é importante que isto seja feito seguindo-se um esquema de categorias da área, caso exista e que pode ser baseado em processos, propriedades, operações, equipamento, etc.

Por último, no estágio de seleção de conceitos pode acontecer que alguns conceitos identificados na etapa anterior sejam descartados, caso se verifique que os mesmos não são úteis para os objetivos a que se destina a indexação da informação.

Vale destacar que existe um grau de complexidade inerente à análise de assunto, a este respeito Naves (2007) destaca que isto se dá por causa do problema de especificação da terminologia e também por causa da subjetividade do indexador, que eventualmente pode se deixar influenciar pelos seus valores individuais.

3 METODOLOGIA

Embora não exista um consenso quanto à metodologia adequada para construção de ontologias (FERNÁNDEZ-LÓPEZ et al, 1997; LUCAS et al, 2008), a análise da literatura mostra que o processo de construção de uma ontologia de domínio deve contemplar as seguintes etapas: aquisição de conhecimento, conceitualização, representação através de uma linguagem formal e avaliação ao longo das etapas de construção da mesma.

Quando se fala em compartilhar conhecimento, fica implícita a necessidade de utilizar conceitos comuns, compreensíveis por toda a comunidade onde este conhecimento é produzido, sendo necessário que estes conceitos que se referem a objetos, agentes, processos, produtos, etc. estejam claramente definidos e sejam amplamente aceitos por esta comunidade. Por outro lado, conforme citado acima, é absolutamente necessário para uma correta modelagem ontológica do domínio que a forma como estes conceitos se relacionam entre si, esteja solidamente estabelecida. Parte-se do pressuposto que quando os objetivos acima enunciados estão corretos e claramente definidos teremos então insumos para a criação de uma ontologia que cubra todos os aspectos necessários para garantir a boa comunicação e troca de conhecimento entre os membros de uma determinada comunidade.

A seqüência das etapas necessárias para se conseguir levantar e compartilhar o conhecimento, que é um dos objetivos de uma modelagem ontológica, pode-se ser sintetizado nas quatro grandes etapas abaixo, as quais, conforme será visto mais adiante subdividem-se em outras:

- 1) Delimitação do domínio que será objeto de estudo;
- 2) Análise de domínio;
- 3) Organização do conhecimento do domínio em uma estrutura facetada;
- 4) Avaliação final do modelo conceitual junto aos especialistas;

A validação do modelo conceitual proposto será realizada em todas as fases, e constitui um procedimento de grande relevância, pois através dele busca-se fazer o ajuste fino da modelagem ontológica.

Embora exista uma ampla tipologia de ontologias (ALMEIDA e BAX, 2003, p.8 a 12), levou-se em consideração que o objetivo da presente pesquisa é focar aspectos relativos à representação do conhecimento, e portanto, trabalhar com ontologias de domínio tornou-se uma escolha natural, uma vez que estas apresentam características que favorecem este propósito.

Em decorrência de ter-se optado por trabalhar com ontologias de domínio e havendo a necessidade de validar a modelagem proposta, o passo seguinte tratou da escolha do domínio de conhecimento a ser estudado e modelado. Visando este objetivo o autor desta pesquisa contactou algumas empresas para as quais foi apresentado o projeto de pesquisa. Desta forma chegou-se até a Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA, por meio da sua unidade voltada às pesquisas de milho e sorgo (<http://www.cnpms.embrapa.br>), localizada na cidade de Sete Lagoas. Esta empresa, através da Diretoria de Pesquisa e Desenvolvimento,

demonstrou bastante interesse nos objetivos de pesquisa apresentados, sendo portanto, no âmbito desta unidade que se desenvolveu a pesquisa aplicada e onde foi efetuada uma parte significativa da coleta de dados.

É importante ressaltar que a EMBRAPA é considerada centro de excelência em pesquisa agropecuária e possui 40 unidades de pesquisa no Brasil, sendo que a unidade de Sete Lagoas é responsável pelo programa de pesquisas das culturas de milho e sorgo.

3.1 Delimitação do domínio

Conforme dito anteriormente, ao estudar o Ser deve-se buscar a melhor maneira de aproximação e de se estabelecer um contato com o mesmo, de forma a permitir conhecê-lo com profundidade. Mas conforme citado por Morin (2007) deve-se ter em mente uma limitação incontestável que fica evidenciada pela incompletude de qualquer conhecimento. Esta busca, no entanto, não deve se eximir de encontrar caminhos que conduzam a uma boa compreensão da essência do Ser e das suas características.

O fato da unidade da EMBRAPA, escolhida para realização da pesquisa aplicada, concentrar as suas atividades de pesquisa nas culturas de milho e sorgo, naturalmente estabeleceu um recorte para este trabalho no que diz respeito ao amplo espectro existente no âmbito das pesquisas agropecuárias. Posteriormente, ao se fazer contato com os pesquisadores da organização, tomou-se conhecimento das especificidades destas duas culturas, o que levou o autor desta pesquisa a limitar o domínio a ser estudado, restringindo-o à cultura do sorgo.

Conforme foi dito na seção 1, foi feita uma análise das metodologias para construção de ontologias amplamente referenciadas na literatura, percebendo-se que das etapas acima citadas, a etapa de aquisição do conhecimento é descrita de forma muito vaga, o que não oferece condições de ter uma visão clara dos procedimentos adotados. No entanto, sabe-se também que o processo de aquisição de conhecimento é uma etapa complexa, pois trata de responder que dizem respeito à delimitação e compreensão da natureza e essência do ser, assim como aspectos que irão definir a abordagem científica mais adequada e que permita estudá-lo melhor.

- **Quem é o Ser?** Como delimito o Ser ou Ente que desejo estudar? De que forma devo cientificamente olhar e me aproximar do Ser?
- **Como captar a essência do Ser?** Como consigo aprender e apreender a essência do Ser?
- **Como capturar os entes formadores de um Ser específico, assim como suas características?**

- **Como modelar a essência do Ser?**

É importante lembrar que o Ser acima citado o qual se pretende modelar emerge do conhecimento que se origina a partir de pesquisas científicas de ponta ligadas à cultura do sorgo. Este corpo de conhecimento envolve questões ligadas não apenas ao cultivo em si, mas a outros aspectos tais como, sistemas de produção, análise de solos, estudo de pragas e doenças, variações genéticas buscando gerar espécies com resistência às condições adversas do clima e solo, etc.

3.2 Levantamento de insumos

Após a etapa de delimitação do domínio, seguiu-se a etapa de levantamento de insumos, o que foi feito buscando identificar e acessar fontes confiáveis e atualizadas que refletissem de forma abrangente o conhecimento teórico e prático desenvolvido e registrado no âmbito das pesquisas agropecuárias focadas na cultura do sorgo. O cumprimento desta etapa conduziu a examinar as fontes citadas a seguir.

3.2.1 Pesquisadores

O primeiro passo foi identificar na própria EMBRAPA, o corpo de pesquisadores que estavam diretamente ligados às pesquisas da cultura do sorgo. A localização destas pessoas-chave permitiu identificar os especialistas do domínio estudado e que em um momento posterior irão exercer função importante na etapa de verificação e avaliação do modelo conceitual desenvolvido neste trabalho de pesquisa. Assim sendo, identificamos este grupo de pessoas a qual é formada por quatro pesquisadores (três seniores e um pleno), que têm apresentado produção científica intensa, sendo que boa parte dos artigos científicos é feita em co-autoria.

Tabela 1 – Produção científica dos pesquisadores da EMBRAPA

PESQUISADOR	TOTAL ARTIGOS	%
José Avelino S. Rodrigues - RODRIGUES, J. A. S.	16	48,5 %
Robert Eugene Schaffert - SCHAFFERT, R. E.	9	27,3 %
Flávio Dessaune Tardin - TARDIN, F. D.	7	21,2 %
Carlos Roberto Casela - CASELA, C. R.	1	3,0 %

Cabe ressaltar que além da EMBRAPA, existem outros centros de pesquisa que dedicam parte das suas atividades à pesquisa do sorgo, entre elas destacam-se:

- A Empresa de Pesquisa Agropecuária de Minas Gerais – EPAMIG é um órgão subordinado à Secretaria de Estado da Agricultura, Pecuária e Abastecimento do governo estadual;
- Departamento de Fitotecnia da Universidade Federal de Viçosa - UFV;
- Departamento de Agricultura da Universidade Federal de Lavras - UFL;
- O Instituto Agrônômico de Campinas - IAC, fundado em 1887, órgão de pesquisa da Agência Paulista de Tecnologia dos Agronegócios, da Secretaria de Agricultura e Abastecimento do Estado de São Paulo, tem sua sede no município de Campinas - SP. A Instituição tem como missão gerar e transferir ciência e tecnologia para o negócio agrícola, visando a otimização dos sistemas de produção vegetal e ao desenvolvimento sócio-econômico com qualidade ambiental.

No entanto é a EMBRAPA Milho e Sorgo a referência nacional em pesquisas ligadas a estas duas culturas. O trabalho aqui apresentado concentrou-se na produção científica deste centro de excelência.

3.2.2 Fontes

A biblioteca da unidade da Embrapa Milho e Sorgo possui no seu acervo diversas pastas que são criadas individualmente para cada pesquisador da instituição.

O objetivo destas pastas é registrar a produção científica de cada pesquisador e nelas são armazenados fisicamente os projetos nos quais houve participação dos pesquisadores, assim como todos os artigos científicos publicados. O acesso a estas pastas que representam fontes riquíssimas e de alta relevância foi disponibilizado para a análise.

3.2.3 Produção literária

O levantamento inicial do material analisado foi realizado considerando as questões levantadas por Barité (2007) referentes à garantia literária e científica, as quais foram detalhadas na seção 2.2. Começando pela questão referente aos documentos que devem ser tomados como referência, a presente pesquisa assumiu que a produção científica dos pesquisadores citados na seção anterior é bastante significativa e representativa do conhecimento produzido nas pesquisas ligadas ao sorgo, tanto nos aspectos literários quanto

científicos. Este conhecimento encontra-se registrado nos artigos científicos e relatórios técnicos que foram localizados na pasta de cada pesquisador.

Outra questão colocada por Barité diz respeito à amostra da literatura (ao longo do tempo) que pode ser considerada significativa, quando se busca validar a garantia literária. Tendo isto em mente, deve-se ressaltar que a produção científica da equipe de pesquisa da EMBRAPA é bastante extensa. Buscando delimitar a nossa amostra e utilizando um critério temporal, após consulta aos pesquisadores, optou-se por considerar para efeitos de amostragem a produção científica compreendida no período de 2006 até 2008, uma vez que este período engloba os interesses de pesquisa mais recentes por parte dos pesquisadores. Tais interesses refletem as demandas pautadas pelo cenário agropecuário brasileiro que, por sua vez, são orientadas por aspectos econômicos do agronegócio e que buscam levar este setor a galgar níveis cada vez mais competitivos no cenário internacional.

Por último, o tipo de conhecimento que se encontra legitimado na literatura selecionada é solidamente ancorado em pesquisas de ponta realizadas nos laboratórios da própria empresa na qual são utilizadas técnicas laboratoriais, procedimentos e metodologias de pesquisas baseados em padrões internacionais e amplamente aceitas no seio da comunidade científica. Outro tipo de conhecimento descrito pela literatura é aquele desenvolvido nos campos experimentais existentes nas próprias dependências da EMBRAPA, onde técnicos agrícolas seguindo rigorosamente especificações fornecidas pelos pesquisadores realizam experiências de campo.

Na etapa de levantamento de dados, baseada na análise da literatura verificou-se no currículo Lattes dos pesquisadores a citação de alguns artigos que não foram localizados fisicamente na pasta dos mesmos. A partir desta constatação e com o intuito de dar maior credibilidade e exatidão à amostra levantada, decidiu-se fazer um cruzamento e verificação dos artigos descritos em ambos os suportes (pastas e currículo Lattes), sendo que em alguns casos só se obteve acesso à integra destes artigos após realizar exaustiva busca, conseguindo localizá-los utilizando mecanismos de buscas da internet. Houve ainda situações em que artigos citados na plataforma Lattes não foram localizados nem nas pastas e nem estavam indexados nos mecanismos de buscas. Quando isto ocorreu foram os próprios pesquisadores que forneceram cópias digitais destes artigos. O levantamento da nossa amostra contendo a produção científica da área ligada ao sorgo, no período de 2006 a 2008, contém um total de 23 artigos.

3.2.4 Veículos de divulgação

A análise dos veículos utilizados pelos membros da comunidade científica ligada ao sorgo para divulgação e publicação das suas pesquisas é um parâmetro significativo, uma vez que é esta própria comunidade que confere ou não credibilidade ao veículo, o que se reflete tanto na qualidade quanto na quantidade da produção científica publicada em cada veículo. Dentro deste critério, foram identificados congressos, publicações e periódicos internacionais que reconhecidamente gozam de alto prestígio no seio desta comunidade.

Foram privilegiados na pesquisa ora apresentada os artigos publicados em periódicos, produzidos por pesquisadores da EMBRAPA, na área do sorgo e que estão relacionados na

Tabela 2 – Veículos de publicação dos pesquisadores da EMBRAPA

VEÍCULO DE DIVULGAÇÃO	ARTIGOS	%
Congresso Nacional de Milho e Sorgo	6	26,1 %
Revista Brasileira de Milho e Sorgo	6	26,1 %
Arquivo Brasileiro de Medicina Veterinária e Zootecnia	6	26,1 %
Congresso Brasileiro de Melhoramento de Plantas	3	13,0 %
Theoretical and Applied Genetics	1	4,3 %
Nature Genetics	1	4,3 %

3.3 Análise dos textos

No seu estudo sobre garantia literária, Barité (2007) sugere refletir sobre os critérios que serão aplicados na análise de documentos e que vão permitir uma correta utilização do critério de garantia literária. A este respeito, utilizou-se como critério o referencial teórico proposto pela análise de assunto e foi através dele que se levou a cabo a análise de conteúdo dos documentos selecionados.

3.3.1 Separar o texto em partes menores

Como ponto de partida para mapear o conhecimento do domínio analisado, foi realizada a leitura na íntegra de cada um dos artigos selecionados. Utilizando um editor de texto, foi criado em paralelo um arquivo e à medida que era feita a leitura do artigo foram sendo separados e transcrevendo neste arquivo trechos do artigo que forneciam informações / conceitos, inicialmente considerados desconhecidos para o autor da pesquisa. Considerou-se

também nesta etapa, conceitos e termos que não fazem parte da linguagem coloquialmente utilizada e que não compõem o senso comum das pessoas.

Os textos selecionados continham termos que se percebeu serem de uso frequente e que constituem o vocabulário corrente usado entre os pesquisadores e especialistas do segmento agropecuário. Deve-se enfatizar que conhecer o vocabulário específico utilizado pelos especialistas ao se comunicarem é de extrema importância pois ajuda a entender a linguagem particular empregada; esta linguagem é composta por termos cujos conceitos carregam intrinsicamente o conhecimento prévio que todo especialista deve possuir.

Segundo Hayes-Roth (1983) o vocabulário inicialmente usado pelo especialista para falar sobre o domínio com um novato é inadequado para a solução de problemas. Isso acontece porque quando o especialista detecta que não está sendo compreendido nos termos triviais (mas específicos da sua área), ele precisa usar termos gerais, que podem ser compreendidos por leigos. Mas esses não são os termos que expressam sua maneira natural de pensar e resolver problemas sobre o domínio. Portanto, qualquer tentativa de criar uma pseudo-linguagem buscando simplificar ou “traduzir” o vocabulário utilizado pelos especialistas carece de legitimidade científica quando o objetivo é retratar com fidelidade o conhecimento do domínio.

Um outro critério utilizado nesta fase foi considerar textos que faziam referência a procedimentos, atividades, processos e índices que forneciam parâmetros numéricos a respeito de variáveis passíveis de mensuração e que são objeto de pesquisa do domínio estudado. Cada texto transposto recebeu um índice numérico seqüencial e também foi identificado em relação ao artigo que lhe deu origem.

3.3.2 Separar texto criando expressões significativas

Em esta etapa foram criadas fichas de enunciados. O processo para preenchimento destas fichas obedeceu ao seguinte critério: a partir dos trechos selecionados na etapa anterior foi realizada uma faturação dos textos, transformando-os em enunciados significativos considerando o escopo do domínio focalizado. Desta forma cada um dos trechos analisados deu origem a vários enunciados ou termos significativos.

Com este procedimento buscou-se rastrear e destacar termos, conceitos, processos e atividades relevantes e que ao mesmo tempo representassem inequivocamente o conhecimento do domínio analisado. Cada ficha criada continha o índice referente à porção do texto donde foi extraído, mantendo também a identificação do artigo que lhe deu origem.

Outro dado importante, contido em cada ficha, diz respeito ao número de ocorrências de cada termo, no total de artigos analisados.

Como forma de ilustrar o processo utilizado segue um exemplo de uma porção do texto selecionado de um artigo e os enunciados e termos que dele foram extraídos.

O sorgo grão tem sido utilizado com intensidade crescente na criação de aves e suínos, principalmente nas regiões produtoras desse grão, como substituto ao milho, por apresentar menor custo de produção ou de aquisição, por encontrar-se mais disponível no mercado nos últimos anos e por apresentar valor nutritivo próximo ao do milho.

O trecho acima deu origem aos seguintes termos: Criação de aves, Criação de suínos, Sorgo grão, Milho, Custo de produção, Custo de aquisição, Valor nutritivo.

Simultaneamente ao processo de seleção de conceitos, deve-se cuidadosamente estabelecer suas conexões, sendo que esta tarefa será realizada fundamentada nas bases teóricas propostas pelas teorias da classificação facetada, com as quais pretende-se construir o sistema de conceitos.

Por último vale destacar que sempre que se julgou necessário, o indexador (autor desta pesquisa) criou termos que embora não constassem nos textos analisados, foram incluídos em fichas por entender que ajudam na compreensão de alguns conceitos. Termos criados segundo este critério foram sinalizados através de colchetes.

4 CONSIDERAÇÕES FINAIS

O passo seguinte a ser desenvolvido e que representa a etapa final a ser cumprida na pesquisa de doutorado diz respeito a avaliação do modelo conceitual por parte dos especialistas do domínio, ou seja, os pesquisadores da EMBRAPA. A avaliação da modelagem proposta inclui a validação da estrutura de termos desenvolvida assim como seus respectivos conceitos, sendo que a validação dos termos permitirá garantir que os mesmos representam o vocabulário especializado utilizado no domínio analisado. A validação dos conceitos referentes a cada termo assim como o relacionamento entre os mesmos, tem como objetivo garantir o arcabouço teórico referente às pesquisas do sorgo, e que por sua vez assegura o compromisso ontológico. Por último, esta etapa de validação garante também na sua essência, que a modelagem proposta reflete o conhecimento dos especialistas, no caso os pesquisadores do sorgo.

MAPPING AND REPRESENTATION OF KNOWLEDGE AND MODELING

Ontological: An experience in the area of sorghum

ABSTRACT

This work stands out the importance of the study of ontologies for Information Science and follows through an analysis of technical literature, several methods developed for ontologies construction. From this analysis and regarding that domain ontologies have as main goal to map the knowledge of a specific area, was checked since the studied literature that this mapping stage which seeks to raise knowledge of specialist, displays few details in general or it's described in a very brief way, making it difficult to understand and use. Based on this observation, the doctoral research reported here assumed the task of filling this gap, developing a method that describes in clear and comprehensive way the stage of knowledge acquisition, and starting from important sources, seeks to structure the knowledge of a domain that will be implemented later in a domain ontology. This method is being used to map the scientific knowledge generated from top research activities developed in farming and cattle sector by EMBRAPA Milho e Sorgo, located at Sete Lagoas city - Minas Gerais.

Key-words: Ontology. Ontologies Construction. Knowledge Acquisition. Farming and Cattle Search.

REFERÊNCIAS

- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**. v. 26, n. 1. p.39-45, set./dez. 2003.
- BARITE, Mario Guido. La Garantia Literaria: Vigencia y Proyección Teórico-Metodológica. **Anais**. VIII ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informação. Salvador - Bahia, Out. 2007.
- CHAUÍ, Marilena. **Convite à Filosofia**. Ed. Ática, São Paulo, 2000.
- DAHLBERG, I. **Ontological structures and universal classification**. Bangalore: Sarada Ranganthan Endowment, 1978. 64p.
- DIAS, E.W; NAVES, M. M. L. **Análise de Assunto**. Brasília: Thesaurus, 2007. 116 p.
- FERNÁNDEZ-LÓPEZ, M.; GÓMES-PÉREZ, A.; JURISTO, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. **Proceedings** of the AAAI97 Spring Symposium Series on Ontological Engineering, March, 1997, p. 33-40.
- FOX, M.; GRUNINGER, M. **Enterprise Modeling**. 1998. Disponível em <http://www.eil.utoronto.ca/enterprise-modelling/papers/fox-aimag98.pdf>. Acessado em fevereiro de 2008.
- FUJITA, Mariângela. A Identificação de Conceitos no Processo de Análise de Assunto para Indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003

GILCHRIST, A. Thesauri, taxonomies and ontologies; an etymological note (2003). EN: **Journal of Documentation** 59(1). 7-18.

HAYES-ROTH, F.; D. A. Waterman; D. B. Lenat. **Building Expert Systems**. 1983, Massachusetts: Addison-Wesley.

MORIN, Edgar. **Introdução ao pensamento complexo**. Lisboa: Instituto Piaget, 2007.

NOY, N. F.; McGUINNESS, D. L. - **Ontology Development 101: A Guide to Creating Your First Ontology**. Disponível em http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html. Acessado em abril de 2006.

QUINE, W. v. O. On What There Is. 1948, in: **From a Logical Point of View**, 2nd revised edition, Harvard University Press: Cambridge, Mass., 1980, pp. 1-19. Disponível em http://en.wikisource.org/wiki/On_What_There_Is. Acessado em 19/12/2008.

SILVA, Daniela Lucas da; SOUZA, Renato Rocha; ALMEIDA, Maurício Barcellos. Comparação de metodologias para construção de ontologias e vocabulários controlados. In: **Anais**. Seminário de Pesquisa em Ontologia no Brasil, 2008, Niterói.

SMITH, B. **Ontology and Informations Systems** (2004). Disponível em <http://www.ontology.buffalo.edu/ontology> Acessado em maio de 2008.

USCHOLD, Mike. GRUNINGER, M. Ontologies: Principles, Methods, and Applications. **Knowledge Engineering Review**, v. 11, n. 2, p. 93-155. June. 1996.

VICKERY, B. C. Ontologies. **Journal of Information Science**, 23 (4) 1997, pp. 277-286. What is Ontology?. Disponível em: <http://www.formalontology.it/index.htm>. Acessado em setembro de 2006.

ANEXO 1

Dissertações e Teses desenvolvidas no Brasil, no escopo da Ciência da Informação, abordando o assunto ontologias.

2009

Viviane N. Pinto de Oliveira	Orientador: Maurício Barcellos Almeida
Dissertação de Mestrado	UFMG - Escola de Ciência da Informação
Uma investigação sobre a avaliação de modelagem conceitual baseada em ontologias: estudo de caso de modelos para sistemas de informação desenvolvidos na Universidade Federal de Minas Gerais	

Júlia Aparecida Gonçalves	Orientador: Renato Souza
----------------------------------	--------------------------

Dissertação de Mestrado	UFMG - Escola de Ciência da Informação
Proposta metodológica para o estabelecimento de relações semânticas em ontologias: contribuição interdisciplinar via Ciência da Informação.	

2008

Daniela Lucas da Silva	Orientador: Renato Souza
Dissertação de Mestrado	UFMG - Escola de Ciência da Informação
Uma comparação de metodologias de construção de tesauros e ontologias.	

2007

Roberto F. Cerqueira	Orientador: Marcello Peixoto Bax
Dissertação de Mestrado	UFMG - Escola de Ciência da Informação
Método de Modelagem Domínio Ontológica do Direito Positivo Brasileiro.	

2006

André Luiz Dias Esbízaro	Orientador: Marisa Brascher Medeiros
Dissertação de Mestrado	UnB - Departamento de Ciência da Informação
Recuperação de informações sobre log de eventos apoiada em ontologia.	

Célia Dias	Orientador: Lídia Alvarenga (em andamento)
Tese de Doutorado	UFMG - Escola de Ciência da Informação
Uso de ontologias no tratamento terminológico de imagens para a melhoria da recuperação da informação.	

Cláudio Gottschalg Duque	Orientador: Marlene de Oliveira
Tese de Doutorado	UFMG - Escola de Ciência da Informação
SIRILICO - Uma proposta para um Sistema de Recuperação de Informação baseado em Teorias da Linguística computacional e Ontologia.	

Edgard Costa Oliveira	Orientador: Mamede Lima-Marques
Tese de Doutorado	UnB - Departamento de Ciência da Informação
Autoria de documentos para a Web Semântica: um ambiente de produção de conhecimento baseado em ontologias.	

Luana Farias Sales	Orientador: Ma. Luiza de A. Campos
---------------------------	------------------------------------

Dissertação de Mestrado	UFF - Ciência da Informação
Ontologias de domínio: estudo das relações conceituais e sua aplicação	
Maurício B. Almeida	Orientador: Ricardo Barbosa
Tese de Doutorado	UFMG - Escola de Ciência da Informação
Um modelo baseado em ontologias para a construção da memória organizacional.	

2005

Marcelo Alvim Jorge	Orientador: Marcello Peixoto Bax
Dissertação de Mestrado	UFMG - Escola de Ciência da Informação
Ontologias no suporte a portais semânticos.	

2003

Alexandra Moreira	Orientador: Lídia Alvarenga
Dissertação de Mestrado	UFMG - Escola de Ciência da Informação
Tesauros e Ontologias: estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o método analítico-sintético.	
Maurício B. Almeida	Orientador: Marcello Peixoto Bax
Dissertação de Mestrado	UFMG - Escola de Ciência da Informação
Inter-operabilidade entre fontes de dados heterogêneas: um meta-modelo baseado em ontologias.	
Sonali Paula Molin Bedinn	Orientador: Lígia Maria Arruda Café
Dissertação de Mestrado	UFSC - Escola de Ciência da Informação
Metodologia para validação de ontologias: o caso ORBIS_MC	