



**FRANCISCO CARLOS DA ROCHA GOMES**

Arquitetura de repositório semântico para organização de pesquisa agropecuária

**Dissertação de mestrado  
Fevereiro de 2013**



**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
ESCOLA DE COMUNICAÇÃO  
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

**FRANCISCO CARLOS DA ROCHA GOMES**

ARQUITETURA DE REPOSITÓRIO SEMÂNTICO PARA ORGANIZAÇÃO DE  
PESQUISA AGROPECUÁRIA

RIO DE JANEIRO  
2013

**FRANCISCO CARLOS DA ROCHA GOMES**

**ARQUITETURA DE REPOSITÓRIO SEMÂNTICO PARA ORGANIZAÇÃO DE  
PESQUISA AGROPECUÁRIA**

Dissertação de Mestrado apresentada  
ao Programa de Pós-Graduação em  
Ciência da Informação, convênio entre  
o Instituto Brasileiro de Informação em  
Ciência e Tecnologia e Universidade  
Federal do Rio de Janeiro / Escola de  
Comunicação, como requisito parcial à  
obtenção do título de Mestre em  
Ciência da Informação

Orientador: Prof. Ph.D. Marcos do Couto Bezerra Cavalcanti

Rio de Janeiro  
2013

G633a

Gomes, Francisco Carlos da Rocha.

Arquitetura de repositório semântico de dados para organização de pesquisa agropecuária / Francisco Carlos da Rocha Gomes. 2013.

169 f. : il. ; 30 cm.

Dissertação (Mestrado em Ciência da Informação) - Programa de Pós- Graduação em Ciência da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Escola de Comunicação, Rio de Janeiro, 2013.

Orientador: Marcos do Couto Bezerra Cavalcanti.

1. Gestão do Conhecimento. 2. Web Semântica. 3. Memória Organizacional. 4. Repositório de dados. 4. Comunicação – Teses. I. Cavalcanti, Marcos do Couto Bezerra. (orient.). II Universidade Federal do Rio de Janeiro, Escola de Comunicação. III Instituto Brasileiro de Informação em Ciência e Tecnologia. IV. Título.

CDU 001.82:630\*26(043.3)

**FRANCISCO CARLOS DA ROCHA GOMES**

**ARQUITETURA DE REPOSITÓRIO SEMÂNTICO PARA ORGANIZAÇÃO DE  
PESQUISA AGROPECUÁRIA**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e Universidade Federal do Rio de Janeiro / Escola de Comunicação, como requisito parcial à obtenção do título de Mestre em Ciência da Informação

Aprovada em 25 de Fevereiro de 2013.

Banca examinadora:

---

Prof. Ph.D. Marcos do Couto Bezerra Cavalcanti (Orientador)  
Programa de Pós-Graduação em Ciência da Informação (IBICT / UFRJ)

---

Prof. Dr. Jorge Calmon de Almeida Biolchini  
Programa de Pós-Graduação em Ciência da Informação (IBICT / UFRJ)

---

Prof. Ph.D. Maria Luiza Machado Campos  
Programa de Pós-Graduação em Informática (UFRJ)

Ao nosso Deus Altíssimo, “porque dele e por ele,  
e para ele, são todas as coisas; glória, pois, a ele  
eternamente. Amém.”  
Romanos 11:36

Aos meus pais, Raimundo Alexandre e Maria da  
Conceição, pelo seu amor e carinho, e por me  
conduzirem pelo caminho dos livros.

## **Agradecimentos**

Agradeço a Deus pela graça de me conduzir até aqui, pois, como disse o profeta Samuel, "Até aqui nos ajudou o SENHOR." 1 Samuel 7:12.

Agradeço à Empresa Brasileira de Pesquisa Agropecuária - Embrapa pela oportunidade de me oferecer condições para cursar este mestrado.

Ao professor Marcos Cavalcante pela orientação neste trabalho e por me apresentar os saberes da complexidade de Morin, por sua longanimidade, dedicação e atenção.

Ao professor Jorge Biolchini por me apresentar os saberes da representação do conhecimento.

À professora Rosali por me apresentar os saberes da organização do conhecimento.

Aos demais professores do PPGCI do IBICT por me apresentarem a Ciência da Informação de uma forma apaixonante e inspiradora.

À minha esposa Rosely e minha filha Juliana, minhas fontes de alegria, inspiração e motivação todos dias, pelo carinho, compreensão.

"Nenhuma mente pode dar um passo sem a ajuda de  
outras mentes."  
Charles S. Peirce, *Semiótica*, p. 48.



## RESUMO

GOMES, Francisco Carlos da Rocha. **Arquitetura de repositório semântico para organização de pesquisa agropecuária**. Orientador: Marcos do Couto Bezerra Cavalcanti. 169 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Comunicação, Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2013.

No âmbito das organizações, a gestão do conhecimento, com seus processos de aquisição, armazenamento e compartilhamento, se baseia em recursos informacionais destinados para o uso direto pelos seres humanos. Porém, estes recursos não são combinantes a ponto de serem usados ou reutilizados para produção de novos conhecimentos emergentes ou inovação. Diante do atual cenário de crescente aumento no volume de informações registradas e do desafio de se compreender cada vez mais e melhor a complexidade dos fenômenos, é preciso contar com alguma forma de memória organizacional integrada e integradora, de maneira atender as necessidades dos usuários em diferentes perspectivas, situações, dimensões, condições e contextos. Por outro lado, a tecnologia da Web Semântica manifesta características que possibilitam o desenvolvimento de uma arquitetura de repositório semântico que atenda essas demandas. Neste sentido, esta dissertação tem como objetivo delinear uma arquitetura de repositório semântico que integre os dados corporativos, podendo servir de referência para implementação em instituição de pesquisa agropecuária.

Palavras-chave: Gestão do Conhecimento. Memória Organizacional.  
Web Semântica.

## ABSTRACT

GOMES, Francisco Carlos da Rocha. **Arquitetura de repositório semântico para organização de pesquisa agropecuária**. Orientador: Marcos do Couto Bezerra Cavalcanti. 169 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Comunicação, Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2013.

Within organizations, knowledge management, with its processes of acquisition, storage and sharing, based on information resources intended for direct use by humans. However, combining these resources are not enough to be used or reused to produce new knowledge or emerging innovation. In today's scenario of increasing the amount of information recorded and the challenge of understanding more and better the complexity of the phenomena, it is necessary to have some form of organizational memory integrated and inclusive, so meet the needs of users in different perspectives, situations, dimensions, conditions and contexts. In contrast, the Semantic Web technology manifests features that enable the development of a semantic repository architecture that meets these demands. In this sense, this paper aims to outline a semantic repository architecture that integrates enterprise data and can serve as a reference implementation in agricultural research institution.

Keywords: Knowledge Management. Organizational Memory. Semantic Web.

## LISTA DE FIGURAS

Figura 1 - Relação triádica entre interpretante, signo e objeto.....	36
Figura 2 - Quatro modos de conversão do conhecimento.....	41
Figura 3 - As quatro operações básicas do conhecimento organizacional.....	44
Figura 4 - Elementos de um sistema de informação.....	54
Figura 5 – Evolução de propriedades rurais.....	62
Figura 6 – Triângulo Semântico.....	63
Figura 7 - Espectro ontológico.....	68
Figura 8 - Interfaces nos sistemas atuais.....	71
Figura 9 - Esquema conceitual como o coração de um sistema integrado.....	72
Figura 10 - Tipos de ontologias de acordo com o nível de dependência.....	75
Figura 11 - Esquema conceitual de integração constituído por ontologias.....	76
Figura 12 - Base de Conhecimento.....	79
Figura 13 - Rede de conceitos para exemplo de inferência semântica.....	80
Figura 14 - Internet Web em comparação com a Web Semântica.....	82
Figura 15 - Pilha da Web Semântica definida pelo W3C.....	85
Figura 16 - Arquitetura de Camadas da Web Semântica.....	86
Figura 17 - Esquema conceitual de triplas RDF.....	90
Figura 18 - Esquema gráfico e respectivo código de tripla RDF.....	90
Figura 19 - Exemplo de definição de vocabulário com RDF/RDFS.....	91
Figura 20 -Vocabulários RDF e FOAF.....	92
Figura 21 - Esquema de classificação utilizando SKOS.....	96
Figura 22 - Arquitetura de integração de dados ligados em RDF.....	101
Figura 23 - Esquema conceitual de anotação de trecho de documento tipo texto .	102
Figura 24 - Construção de ontologias Methontology.....	104
Figura 25 - Arquitetura de repositório semântico.....	114
Figura 26 - Mapa Conceitual da estratégia metodológica da dissertação.....	138
Figura 27 - Mapa conceitual das relações semânticas.....	139
Figura 28 - Mapa conceitual do processo de conceitualização de Dahlberg.....	140
Figura 29 - Mapa conceitual da definição de um conceito.....	141

Figura 30 - Mapa conceitual sobre a análise de intenção e extensão.....	142
Figura 31- Mapa conceitual do processo de síntese de um conceito.....	143
Figura 32 - Mapa conceitual das premissas para conceitualização.....	144
Figura 33 - Mapa conceitual da relações entre conceitos.....	145
Figura 34 - Mapa conceitual da linguagem de marcação SGML.....	146
Figura 35 - Uso de RDF e FOAF e listagem de triplas.....	147
Figura 36 - Áreas de conhecimento da Embrapa Acre.....	150
Figura 37 - Áreas de pesquisa da Embrapa Acre.....	151
Figura 38 - Esquema de procedimento científico de Quivy e Campenhoudt.....	152
Figura 39 - Página na Internet do programa CmapTools.....	153
Figura 40 - Página de download do programa CmapTools.....	154
Figura 41 - Mecanismo de busca de termos no CmapTools.....	155
Figura 42 - Referências e histórico do vocabulário SKOS.....	157
Figura 43 - Referências e histórico do vocabulário Dublin Core.....	157
Figura 44 - Lista de vocabulários SKOS.....	158
Figura 45 - Página do portal na Internet do Tesouro Agrovoc da FAO.....	159
Figura 46 - Termo Embrapa na DBPedia.....	160
Figura 47 - Termo Cereal na DBPedia.....	161
Figura 48 - Termo Cassava (mandioca) na DBPedia.....	162
Figura 49 - Página de validação RDF do W3C.....	163
Figura 50 - Resultado da validação de RDF.....	164
Figura 51 - Organograma da Embrapa.....	165
Figura 52 - Unidades da Embrapa no território nacional.....	166
Figura 53 - Página do AGROVOC sobre milho.....	167
Figura 54 - Parte final da página do AGROVOC sobre milho.....	168
Figura 55 - Página da DBPedia sobre milho.....	169

## LISTA DE ABREVIATURAS E SIGLAS

CERN	<i>European Organization for Nuclear Research</i> , Organização Européia para Pesquisa Nuclear
CMS	<i>Content Management Systems</i> (Sistema de Gestão de conteúdo)
DL	<i>Description Logic</i> (Lógica Descritiva)
DM	<i>Data Mining</i> (Mineração de Dados)
DTD	<i>Document Type Definition</i> (Definição de Tipo de Documento)
DW	<i>Data Warehouse</i> (Armazém de Dados)
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EP	Emergent patterns (Padrões emergentes)
FAO	<i>Food and Agriculture Organization of the United Nations</i> (Organização das Nações Unidas para Agricultura e Alimentação)
HTML	<i>HyperText Markup Language</i> (Linguagem de Marcação de Hipertexto)
HTTP	<i>Hiper Text Transfer Protocol</i> (Protocolo de Transferência de Hipertexto)
IA	Inteligência Artificial
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IRI	<i>Internationalized Resource Identifier</i> (Identificador de Recurso Internacionalizado)
KB	<i>Knowledge-base</i> (Base de conhecimento)
KDD	<i>Knowledge Discovery in Databases</i> (Descoberta de Conhecimento em Bancos de Dados)
KO	<i>Knowledge Organization</i> (Organização do Conhecimento)
KOS	<i>Knowledge Organization Systems</i> (Sistemas de Organização do Conhecimento)
OC	Organização do Conhecimento
OWL	<i>Web Ontology Language</i> (Linguagem de Ontologia da Web)
RDF	<i>Resource Description Framework</i> (Conjunto de Classes de Descrição de recursos)
RDFS	<i>Resource Description Framework Schema</i> (Esquema de Conjunto de Classes de Descrição de Recursos)
RC	Representação do conhecimento
RIF	<i>Rule Interchange Format</i> (regra de formato de intercâmbio)
SGML	<i>Standard Generalized Markup Language</i> (Padrão Generalizado de Linguagem de Marcação)
SKOS	<i>Simple Knowledge Organization Systems</i> (Sistemas Simples de Organização do Conhecimento)
SPARQL	<i>SPARQL Protocol and RDF Query Language</i> (acrônimo recursivo de Protocolo e Linguagem de Consulta para RDF)
TICs	Tecnologias de Informação e Comunicação
URI	<i>Uniform resource identifier</i> (Identificador Uniforme de Recurso)
URL	<i>Unified Resource Locator</i> (Unidade de Localização de Recursos)
W3C	<i>World Wide Web Consortium</i>
WS	Web Semântica
WWW	<i>World Wide Web</i> (grande teia mundial da Internet)
XML	<i>eXtensible Markup Language</i> (Linguagem de Marcação Extensível)

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>17</b>
<b>1.1 Escopo da dissertação.....</b>	<b>19</b>
<b>1.2 Contexto do problema.....</b>	<b>19</b>
<b>1.3 Objetivos.....</b>	<b>21</b>
<b>1.4 Metodologia.....</b>	<b>22</b>
1.4.1 Fundamentos teóricos da metodologia.....	22
1.4.2 Aplicação metodológica.....	25
<b>2 MARCOS CONCEITUAIS E TEÓRICOS.....</b>	<b>28</b>
<b>2.1 Crescimento e complexidade do conhecimento.....</b>	<b>28</b>
<b>2.2 Dado, informação e conhecimento.....</b>	<b>34</b>
2.2.1 Dado.....	34
2.2.2 Informação.....	37
2.2.3 Conhecimento.....	39
2.2.3.1 <i>Conhecimento tácito</i> .....	41
2.2.3.2 <i>Conhecimento explícito</i> .....	42
<b>2.3 Conhecimento organizacional e gestão do conhecimento.....</b>	<b>43</b>
<b>2.4 A memória das organizações.....</b>	<b>46</b>
<b>2.5 A gestão de documentos.....</b>	<b>48</b>
<b>2.6 Organização do conhecimento.....</b>	<b>53</b>
<b>2.7 Representação do conhecimento.....</b>	<b>60</b>
<b>2.8 Sistema de informação baseado em ontologias.....</b>	<b>70</b>
<b>2.9 Web semântica.....</b>	<b>82</b>
2.9.1 A arquitetura da Web Semântica .....	83
2.9.2 Resource Description Framework (RDF) e RDF Schema.....	88
2.9.3 OWL – Linguagem para Construção de Ontologias na Web.....	93
2.9.4 Dados ligados (Linked Data).....	96

<b>3 ARQUITETURA DE REPOSITÓRIO SEMÂNTICO DE DADOS.....</b>	<b>98</b>
<b>3.1 Arquitetura de repositório semântico.....</b>	<b>106</b>
3.1.1 Organização e representação do conhecimento da arquitetura.....	109
3.1.2 A arquitetura.....	111
3.1.3 Processo de desenvolvimento e manutenção da arquitetura.....	115
<b>4 CONCLUSÃO.....</b>	<b>117</b>
<b>4.1 Considerações finais.....</b>	<b>117</b>
<b>4.2 Perspectivas para trabalhos futuros.....</b>	<b>120</b>
<b>REFERÊNCIAS.....</b>	<b>122</b>
<b>APÊNDICES.....</b>	<b>138</b>
Apêndice A - Mapa conceitual do processo metodológico desta dissertação.....	138
Apêndice B - Mapa conceitual das relações semânticas.....	139
Apêndice C - Mapa conceitual do processo de conceitualização.....	140
Apêndice D - Mapa conceitual da definição de um conceito.....	141
Apêndice E - Mapa conceitual sobre a análise de intenção e extensão.....	142
Apêndice F - Síntese das características de um conceito - análise intensional.....	143
Apêndice G - Premissas para conceitualização – análise intensional.....	144
Apêndice H - Relações entre conceitos para análise extensional.....	145
Apêndice I - Mapa conceitual da linguagem genérica de marcação SGML.....	146
Apêndice J - Exemplo de uso de vocabulários RDF e FOAF.....	147
Apêndice K - Códigos em RDF/XML.....	148
Apêndice L - Áreas do conhecimento da Embrapa Acre.....	150
Apêndice M - Áreas de Pesquisa da Embrapa Acre.....	151
<b>ANEXOS.....</b>	<b>152</b>
Anexo A - Esquema de procedimento científico.....	152
Anexo B - Programa de Computador CmapTools.....	153
Anexo C - Vocabulários controlados da web semântica.....	156
Anexo D - Vocabulários SKOS.....	158

Anexo E - Termos da DBpedia.....	160
Anexo F - Validação de RDF.....	163
Anexo G - Organograma da Embrapa.....	165
Anexo H - Unidades da Embrapa no território nacional.....	166
Anexo I - Conexão do tesaouro Agrovoc com a DBpedia.....	167



## 1 INTRODUÇÃO

Na organizações de pesquisa científica, que buscam constantemente de novos conhecimentos e inovação, a informação é como o sangue. E como sangue significa vida, a informação é vida que flui e promove crescimento do conhecimento na organização.

Quando o sangue não flui de maneira correta, surge debilidade ou doença. Da mesma forma, informação que não flui adequadamente pela organização gera um quadro patológico: perda de informação, informação disponível mas insuficiente e informação fragmentada.

Os problemas informacionais são de natureza complexa, com causas e efeitos se combinando, muitas vezes sem uma relação direta. Aspectos estáticos podem não estar em harmonia com os aspectos dinâmicos. Por exemplo, estruturas de informação podem ser elementos que, com o tempo, não satisfazem mais as necessidades dos agentes.

Os problemas tornam-se imperceptíveis, difíceis de se identificar e compreender. A perda ou insuficiência de informação tornam-se como pequenos e constantes vazamentos que levam, em conjunto, a um grande desperdício ou carência de recursos, comprometendo os processos de geração de conhecimento e inovação da organização. Isto pode ocorrer mesmo se, aparentemente, as condições de operação e resultados esperados sejam satisfatórios.

Mesmo diante de um quadro aparentemente satisfatório, como é possível alcançar um nível superior de excelência e de novas possibilidades de inovação a partir do conhecimento gerado na organização?

Um passo inicial pode ser o entendimento de que as soluções informacionais não são primordialmente tecnológicas.

Pessoas nas organizações fazem uso de tecnologia para resolver seus problemas informacionais. Porém, estes não são de natureza tecnológica, mas constituídos de idiosincrasia própria, exigindo gestão e visão holista específicas sobre a complexidade dos seus fenômenos.

Gestão da informação se confunde com gestão da tecnologia da informação

em virtude das soluções informacionais utilizarem amplamente algum tipo tecnologia. Neste sentido, a tecnologia se destaca mais, torna-se mais evidente, apresentando-se como principal elemento solucionador dos problemas informacionais.

As tecnologias são transitórias, mudam e se aperfeiçoam cada vez mais a fim de resolver efetivamente os problemas informacionais, cada vez mais complexos. Por outro lado, a solução destes passa fundamentalmente pela análise de sua essência, sua natureza, causas e efeitos, resultando em estratégias e abordagens específicas, de onde destacam-se a organização e representação do conhecimento, que possibilita a fluidez e a memória da informação e do conhecimento nas diversas instâncias componentes do contexto corporativo.

Numa organização, é importante considerar a necessidade de tornar explícito o conhecimento que está na mente das pessoas, a fim de retê-lo, aprimorá-lo e compartilhá-lo, formando uma memória organizacional que conduz a um salto qualitativo maior de conhecimento, fundamental para um ambiente de geração de inovação.

Sem memória a organização pode não ter como renovar e fortalecer seu conhecimento corporativo. Berners-Lee (1989) lembra da alta rotatividade de pessoas no CERN (*European Organization for Nuclear Research*, Organização Européia para Pesquisa Nuclear) e como isto prejudicava o andamento dos trabalhos científicos, dada à quebra do fluxo de conhecimento tácito (na mente) para o explícito (externo), com muitos conhecimentos não compartilhados e perdidos.

É preciso olhar os problemas na perspectiva holística da complexidade dos fenômenos e da gestão do conhecimento para, somente depois, adotar uma solução tecnológica mais adequada.

Uma visão holística permite perceber tanto a essência dos problemas, quanto suas causas e efeitos, que se relacionam de forma distante e indireta. Segundo Kahane:

Os problemas de alta complexidade só podem ser resolvidos por meio de processos sistêmicos [...] Um problema tem uma baixa complexidade 'dinâmica' se a causa e o efeito estiverem próximos no tempo e no espaço [...] as causas produzem efeitos que estão próximos, são imediatos e óbvios. [...] Um problema tem uma alta complexidade dinâmica se a causa e

o efeito estiverem distantes no espaço e no tempo [...] Tais problemas só podem ser entendidos de forma sistêmica, levando-se em conta as inter-relações de suas partes e o funcionamento do sistema como um todo. (KAHANE, 2008, p. 64-66)

Neste contexto, o papel da tecnologia é de ser suporte fundamental para lidar com a complexidade e a dinâmica dos processos, bem como o volume cada vez maiores das informações das organizações.

## **1.1 Escopo da dissertação**

Esta dissertação trata de questões relativas à gestão do conhecimento explícito utilizando Web Semântica para aplicação em organização de pesquisa científica no campo da agropecuária.

Vale ressaltar que este trabalho não discute questões tecnológicas. Restringe-se apenas à sua aplicação no campo da gestão do conhecimento.

## **1.2 Contexto do problema**

O conhecimento tem se tornado um fator cada vez mais relevante para o desenvolvimento da sociedade. A experiência acumulada sobre a terra e os demais recursos naturais foi importante para a promoção da agricultura e o estabelecimento das relações mercantis. O progresso e aperfeiçoamento do conhecimento técnico possibilitou o desenvolvimento de ferramentas e máquinas que imitam, estendem e potencializam as capacidades e habilidades humanas, resultando, dentre outros benefícios, na produção de bens em larga escala a custos menores, tornando sua aquisição cada vez mais acessível a toda sociedade.

Na atualidade, os esforços da sociedade se direcionam de uma economia orientada por produtos e serviços para uma economia guiada pelo conhecimento e inovação (SCHARMER, 2010, p. 59). O progresso técnico das tecnologias de informação e comunicação (TICs), possibilitou amplo uso da informação com benefícios cada vez mais evidentes em todas as áreas de atuação da sociedade.

Estes avanços são em função do emprego cada vez maior do conhecimento científico, que se distingue do conhecimento comum (popular) por sua confiabilidade e submissão à avaliação, validação e consenso da comunidade (MUELLER, 2000, p. 21; 2006, p. 28).

Segundo Le Coadic:

As atividades científicas e técnicas são o manancial de onde fluem os conhecimentos científicos e técnicos que se transformarão, depois de registrados, em informações científicas e técnicas [...] Ademais, o processamento desses conhecimentos, que se torna possível após entrarem em circulação, está na origem das descobertas científicas e das inovações técnicas (LE COADIC, 2004, p. 26).

No progresso científico atual destacam-se dois desafios:

- a) o crescimento da informação e das necessidades de conhecimento: apesar do emprego cada vez maior das TICs para gerenciar o registro, o armazenamento e a disponibilidade do crescente volume de informação, a capacidade de acesso e de assimilação por parte dos usuários tornam-se desafios;
- b) a crescente complexidade dos problemas científicos, que necessitam cada vez mais de uma abordagem sistêmica, holística, complexa, além de conhecimento emergente.

As demandas de pesquisas científicas são cada vez mais crescentes por gestão de dados dinâmicos e heterogêneos (KIRYAKOV; DAMOVA, 2011, p. 236). Bases de conhecimentos científicos exigem representação do conhecimento mais elaborada, que apreenda melhor os fenômenos, as dimensões e as dinâmicas em que se inserem.

No âmbito organizacional diversos autores elencam problemas latentes de gestão da informação (OGBUJI et al., 2007; DOLBEAR, 2007):

- a) coleta e armazenamento fragmentado de dados, ocasionando a segmentação da informação;
- b) bases de dados dispersas ou isoladas, destinadas a fins específicos, sem nenhuma ou com pouca integração com outras bases da própria organização;

- c) dados e informação representados por meio de terminologia ambígua, idiossincrática, sem padronização definida;
- d) soluções corporativas de gestão da informação mais alinhadas com repositórios de documentos e bases de dados do que as bases de conhecimento;
- e) ausência de normas e procedimentos que permitam o estabelecimento e uso de memória institucional de base de conhecimento, orientada para a geração de conhecimento e inovação.

Por outro lado, repositórios semânticos de dados baseados na tecnologia de Web Semântica (WS) apresentam características que podem solucionar muitos desses problemas. Eles possuem uma forma padrão e escalável de representar a informação, facilitando a integração de diferentes fontes de dados. Com Web Semântica as bases de dados das organizações podem tornar-se bases de geração de novos conhecimentos, pois permite agregar significado aos dados mais elementares, possibilitando melhor combinação e recuperação das informações, bem como processamento de inferências computacionais.

### **1.3 Objetivos**

O presente trabalho pretende contribuir com o delineamento de uma arquitetura de repositório semântico de dados científicos com tecnologia de Web Semântica.

Este trabalho tem como objetivo geral avaliar o papel da arquitetura de repositório de dados com Web Semântica para a gestão do conhecimento corporativo no âmbito da organização de pesquisa científica no campo da agropecuária.

Além disso, serão desenvolvidos os seguintes objetivos específicos:

- a) estudar a natureza do conhecimento científico na perspectiva do processo

de geração, retenção e compartilhamento do conhecimento, dentro do contexto das crescentes demandas de pesquisa e suas necessidades de visão complexa, holística, e sistêmica;

- b) mapear elementos que caracterizam uma arquitetura de repositório semântico de dados com a finalidade de promover o compartilhamento de conhecimento semanticamente estruturado em nível de dados, possibilitando o processamento de inferências, e facilitar a operação e intercomunicação entre sistemas de informação a partir do estabelecimento de uma terminologia comum no âmbito de uma organização de pesquisa agropecuária;
- c) adequar o modelo genérico da arquitetura às especificidades da organização de pesquisa científica no campo da agropecuária, Embrapa.

## **1.4 Metodologia**

Para alcançar os objetivos, foi adotada a pesquisa qualitativa, exploratória e descritiva, por método dedutivo baseado em conceitualizações, considerando a literatura existente como fonte de informação.

### **1.4.1 Fundamentos teóricos da metodologia**

Segundo Pedro Demo, a pesquisa qualitativa trata dos aspectos essenciais<sup>1</sup> do objeto de estudo (DEMO, 2000, p. 146) e não busca o "dado" propriamente dito, mas a "informação discutida", produzida por meio de interações (DEMO, 2011, p. 108).

A pesquisa qualitativa limita-se a descrever, analisar e classificar fatos sem a interferência do pesquisador. O objeto de pesquisa pode ser bibliográfico, possibilitando a construção de panoramas sobre determinado assunto (SANTAELLA,

---

1 A palavra 'qualidade' vem da palavra '*qualitas*', que significa 'essência' em latim.

2001, p. 146-147).

A pesquisa qualitativa envolve as atividades de observação, registro, análise e correlação de fatos ou fenômenos sem manipulá-los. Busca conhecer as diversas situações e relações do fenômeno estudado. E pode assumir a forma de estudo exploratório, sem a necessidade da elaboração de hipóteses a serem testadas no trabalho, restringindo-se em buscar maiores informações e como elas se relacionam entre si no âmbito de determinado assunto, visando a familiaridade com o fenômeno ou obter nova percepção do mesmo e descobrir novas ideias (CERVO; BERVIAN, 1996, p. 49).

A pesquisa exploratória e descritiva procura conhecer trabalhos anteriores por meio da literatura correlata que, com releituras, conduz à reflexão, à compreensão de novas relações entre os fatos, proporcionando condições para o surgimento de novos significados e novas compreensões para os fenômenos estudados (QUIVY; CAMPENOUDT, 1995, p. 49, 85 ; GIL, 1999, p. 44).

Sobre pesquisa bibliográfica, Marconi e Lakatos (2003, p. 183) afirmam que esta “não é mera repetição do que já foi dito ou escrito sobre certo assunto, mas propicia o exame de um tema sob novo enfoque ou abordagem, chegando a conclusões inovadoras.”

A pesquisa por método dedutivo envolve questionamentos e identificação de conceitos dentro de um domínio de conhecimento, que podem servir de base para o desenvolvimento de novas questões e conceitos, bem como resultar em novas conclusões, num processo cíclico de consolidação do conhecimento adquirido e descoberta de novos conhecimentos e significados.

Popper detalha o processo dedutivo da seguinte forma:

A partir de uma ideia nova, formulada conjecturalmente e ainda não justificada de algum modo - antecipação, hipótese, sistema teórico ou algo análogo - podemos tirar conclusões por meio de dedução lógica. Essas conclusões são em seguida comparadas entre si e com outros enunciados pertinentes, de modo a descobrir-se que relações lógicas (equivalência, dedutibilidade, compatibilidade ou incompatibilidade) existem no caso. (POPPER, 1975, p. 33)

O método dedutivo parte de um entendimento genérico e verdadeiro afim de verificar sua aplicabilidade a casos específicos (GIL, 1999, p. 27; HYDE, 2000, p.

83). Na abordagem dedutiva, conclusões específicas se baseiam em premissas gerais (WHEELDON; AHLBERG, 2012, p. 114).

Além disso, segundo Quivy e Campenhoudt, a elaboração de conceitos (conceitualização) é importante para a construção do modelo de análise da pesquisa, de forma que este não seja vago, impreciso e arbitrário (QUIVY; CAMPENOUDT, 1995, p. 111). Para estes autores, existem duas formas de construir conceitos: por meio da indução, que gera conceitos operatórios, isolados; e por meio da dedução, que cria conceitos sistemáticos (op. cit., p. 122-123). E quanto a estes últimos, os autores destacam suas vantagens em relação aos primeiros:

O conceito sistêmico não é induzido pela experiência; é construído por raciocínio abstrato – dedução, analogia, oposição, implicação, etc. -, ainda que se inspire forçosamente no comportamento dos objetos reais e nos conhecimentos anteriormente adquiridos acerca desses objetos. (op. cit., p. 125)

Quivy e Campenhoudt assim resumem o processo de conceitualização por dedução: “Situamos o conceito em relação a outros conceitos e, depois, por meio de deduções em cadeia, isolamos as dimensões, as componentes e os indicadores” (op. cit., p. 135).

O processo de conceitualização também é abordado por David Ausubel em sua teoria cognitiva de aprendizagem significativa, que assim é descrita por Moreira:

A aprendizagem é dita significativa quando uma nova informação (conceito, ideia, proposição) adquire significados para o aprendiz através de uma espécie de ancoragem em aspectos relevantes da estrutura cognitiva preexistente do indivíduo, isto é, em conceitos, ideias, proposições já existentes em sua estrutura de conhecimentos (ou de significados) com determinado grau de clareza, estabilidade e diferenciação ... Na aprendizagem significativa há uma interação entre o novo conhecimento e o já existente, na qual ambos se modificam. À medida que o conhecimento prévio serve de base para a atribuição de significados à nova informação, ele também se modifica. ” (MOREIRA, 1998, p. 5)

Com base nesta teoria, Joseph Novak concluiu que mensagens escritas ou faladas são seqüências de conceitos e proposições e que o conhecimento armazenado em nossas mentes se dispõe em estruturas hierárquicas ou holográficas, em arranjo não-linear (NOVAK; GOWIN, 1984, p. 53).

Desta conclusão Novak (1998, p. 23) desenvolveu o mapa conceitual: uma



técnica que permite construir estruturas organizadas de conhecimento por meio da representação de conceitos, bem como relações de significância entre conceitos na forma de proposições.

Um mapa conceitual é um grafo, um conjunto de conceitos (vértices) ligados entre si (arestas), onde tais ligações possibilitam atribuir significado ao conjunto dos conceitos envolvidos nas ligações.

Para este trabalho foi adotado o programa de computador *CmapTools*<sup>2</sup>, que agiliza a elaboração e integração de mapas conceituais, e que possibilita a conexão de cada conceito com *links* da Internet e arquivos contendo textos, imagens ou qualquer outro tipo de dado.

Destaca-se, como recurso que aumenta a produtividade, o mecanismo de busca do *software CmapTools* para localização rápida de termos em diferentes mapas conceituais. É um recurso útil para manter a consistência entre diversos mapas, pois ao longo do processo de dedução e elaboração dos mapas conceituais, um mesmo termo pode estar definido em diversos mapas. Porém, com este recurso de busca, é possível localizar e revisar os termos de maneira que cada termo tenha as mesmas características em todos os mapas em que ocorre.

O processo metodológico de conceitualização por dedução para este trabalho encontra-se disposto em mapa conceitual no Apêndice A. Enquanto que detalhes do *software CmapTools* encontram-se no Anexo B.

Este trabalho também se baseia no esquema de procedimento científico definido por Quivy e Campenhoudt (1995, p. 27) para a condução de pesquisas em ciência sociais, cujo modelo está descrito no anexo A.

## 1.4.2 Aplicação metodológica

Para o desenvolvimento da dissertação foram utilizadas as sete etapas do esquema de procedimento científico definido por Quivy e Campenhoudt (1995, p. 27) para a condução de pesquisas em ciência sociais (Anexo A):

---

2 Disponível em: <http://cmap.ihmc.us/>. Acessado em: 2 fev. 2013.

- 1) pergunta de partida;
- 2) exploração das leituras;
- 3) estudo da problemática;
- 4) a construção do modelo de análise com mapas conceituais;
- 5) observação;
- 6) análise das informações;
- 7) conclusões.

Na primeira etapa, a pergunta de partida para a pesquisa foi: **como a Web Semântica pode contribuir para melhoria da gestão do conhecimento de uma empresa de pesquisa agropecuária?**

Na segunda e terceira etapas, para a pesquisa bibliográfica e estudo da problemática, buscou-se na produção científica os trabalhos referentes aos estudos sobre:

- a) conceitos fundamentais de gestão do conhecimento, gestão da informação, organização e representação da informação;
- b) a natureza da complexidade e visão holística sistêmica do conhecimento científico;
- c) a tecnologia de web semântica e suas potencialidades;
- d) memória organizacional e a arquitetura de repositório semântico.

Nas etapas quatro, cinco e seis - construção do modelo de análise, observação e análise das informações, foi utilizado o método analítico-sintético de análise dos conceitos e síntese de suas características, conforme Dahlberg (1978) (Apêndice C).

Para a última etapa, conclusão, foi feito o delineamento da arquitetura e repositório semântico, tendo como base de referência:

- a) a arquitetura de integração de dados com Web Semântica proposta por Berners-Lee (2005);
- b) o método de construção de ontologias Methontology (GÓMEZ-PÉREZ et al., 2004, p. 135);
- c) as definições de ontologia de domínio para empresa e gestão do conhecimento (GÓMEZ-PÉREZ et al., 2004, p. 85, 98, 102);

- d) a estrutura orgânica da Embrapa, tanto em âmbito nacional, quanto no contexto da Unidade Descentralizada da Embrapa Acre, procurando identificar as atividades institucionais, áreas de conhecimento em que atua, agentes de informação e suas necessidades, as fontes de informação e seus fluxos;
- e) consulta e análise do conteúdo do Regime Interno da Embrapa Acre e das páginas da Embrapa Sede e Embrapa Acre, afim de identificar estruturas de informação, conceitos e suas relações.

## 2 MARCOS CONCEITUAIS E TEÓRICOS

### 2.1 Crescimento e complexidade do conhecimento

O ser humano é capaz de saber mais do que pode expressar (POLANYI, 1983, p. 18). Porém, com o aumento da eficiência do registro, armazenamento e recuperação das informações proporcionado pelas TICs, a geração de informação se desenvolve numa escala maior do que podemos assimilar (SWANSON, 2001, p. 12).

Segundo Frawley et al.:

[...] existe uma crescente distância entre a geração de dados e de compreensão de dados. Ao mesmo tempo, existe um crescente percepção e expectativa de que os dados, analisados e apresentados de maneira inteligente, será uma fonte valiosa para ser usado para uma vantagem competitiva. (FRAWLEY, 1992, p. 57)

Por outro lado, o avanço e a expansão do conhecimento científico exige cada vez mais grandes volumes de informações de qualidade (bem organizadas e confiáveis), com a finalidade de satisfazer suas demandas.

Para Vannervar Bush (1945, p. 2-3), “um registro, se for para ser útil à ciência, deve ser continuamente estendido, deve ser armazenado, e acima de tudo, deve ser consultado”. E “a soma da experiência humana está a se expandido a uma taxa prodigiosa”.

Popper afirma que o conhecimento crescente não é uma simples sequência de sistemas dedutivos cada vez mais aperfeiçoados, mas sim o avanço de velhos problemas para novos problemas por meios de conjecturas e refutações, formando uma espécie de “árvore do conhecimento” ou “árvore evolucionária” que, “sob a influência de problemas cada vez mais especializados, ramificou-se em vasto número de formas altamente especializadas.” (POPPER, 1960, p. 176; 1999, p. 236-239)

A expansão científica, somada à alta capacidade de processamento das TICs, gera grande quantidade de dados de difícil gestão:

“Instrumentos científicos e simulações de computador estão criando arquivos de grandes extensões que exigem novos métodos científicos para analisar e organizar os dados. Volumes de dados estão aproximadamente dobrando a cada ano. Uma vez que novos instrumentos têm precisão extraordinária, a qualidade dos dados tendem a melhorar rapidamente. A análise destes dados para encontrar os efeitos sutis que passam despercebidos em estudos anteriores requer algoritmos que podem, simultaneamente, lidar com grandes conjuntos de dados e que podem encontrar efeitos muito sutis - encontrando as agulhas no palheiro, ou seja, informações muito sutis que não foram detectadas em medições anteriores.” (GRAY, 2005, p. 34)

A geração de conhecimento é um processo social, complexo, dinâmico, não linear, dependente de um ambiente adequado. O ambiente adequado é caracterizado por ações de colaboração, cooperação, solidariedade, mutualidade, transparência, corresponsabilidade, participação dos atores, troca de experiências, discussão de problemas comuns, compartilhamento de soluções, e integração das pessoas.

Hinton (2009, p. 37-39) considera o contexto como característica das dimensões de tempo e espaço. O conhecimento compartilhado, consensual, compreensivo de forma homogênea em diversos ambientes compreende o contexto espacial. Por outro lado, com o tempo, também podem ocorrer as mudanças nos significados das informações.

O contexto, segundo Scharmer (2010, p. 103-131), é um “laboratório vivo”, o lugar no qual a atividade da observação é realizada, e para onde nosso campo de atenção se desloca com o objetivo de se obter uma compreensão mais abrangente. E, acrescenta: “o conhecimento original é fruto de um todo interligado, em vez de partes contingentes isoladas”.

Segundo Morin e Le Moigne (2009, p. 36), as formas de criação de conhecimento pela ciência clássica se baseiam em três princípios fundamentais:

- 1) o determinismo universal, que possibilita conhecer passado e futuro apenas com desenvolvimento da inteligência e dos sentidos;
- 2) o reducionismo, onde um todo pode ser explicado pela análise de uma parte;

3) e a disjunção, onde a apreensão do conhecimento dar-se-á pelo método de divisão e conquista, ou seja, para se compreender melhor o todo é preciso isolar e separar o objeto de conhecimento em partes.

Anderson afirma que o pensamento reducionista, do qual se pode reduzir tudo a simples leis fundamentais, encontra dificuldades de explicar certos fenômenos por causa de aspectos de escala e complexidade. E complementa: "Na verdade, quanto mais a física de partículas elementares nos fala sobre a natureza das leis fundamentais, menos relevância elas parecem ter para os problemas reais do resto da ciência, muito menos aos da sociedade." (ANDERSON, 1972, p. 393)

Para Morin e Le Moigne (2009, p. 29, 38, 44), as antigas formas de criação de conhecimento já não são suficientes para garantir a sustentabilidade do processo generativo de novas ideias a partir da realidade que se apresenta cada vez mais complicada e complexa. Apenas o entendimento da ordem e leis universais (determinismo) não garantem a real compreensão da realidade que apresentam aspectos como desordem, dispersão e desintegração.

Complexidade vem do latim *complexus*, que significa "o que é tecido conjuntamente". É um fenômeno que se caracteriza por um número muito grande elementos que se interagem e se interferem mutuamente. (MORIN, 2011, p. 35)

Para Scharmer (2010, p. 45-46), a complexidade possui um aspecto dinâmico, onde a relação de causa e efeito é constituída por uma cadeia que pode ser longa e mais interdependente, envolvendo muitos elementos, podendo cada um deles agir de forma distinta como, por exemplo, no caso da complexidade social, onde agentes envolvidos podem agir conforme seus interesses, de maneira independente.

Segundo Kahane, para resolver um problema complexo é preciso mergulhar na sua complexidade, considerando e refletindo tanto sobre o que é concordante quanto o que é discordante em relação ao problema, avançando além do conhecimento passado. E acrescenta:

Quando falamos em 'resolver um problema', partimos do princípio de que estamos separados do problema e que podemos estudá-lo de forma objetiva e controlá-lo mecanicamente, com a causa produzindo o efeito, como no

caso de um carro quebrado. Mas esse não é um bom modelo do nosso mundo, cada vez mais complexo e interdependente, de rápidas mudanças. Não há 'um' problema ao qual podemos reagir e consertar. Há uma 'situação-problema' da qual cada um de nós é parte, assim como um órgão é parte do corpo. (KAHANE, 2008, p. 157)

Segundo Capra, o conhecimento fragmentado é consequência do método analítico, e com tendência ao reducionismo:

O método de Descartes é analítico. Consiste em decompor pensamentos e problemas em suas partes componentes e em dispô-las em sua ordem lógica. Esse método analítico de raciocínio é provavelmente a maior contribuição de Descartes à ciência. Tornou-se uma característica essencial do moderno pensamento científico e provou ser extremamente útil no desenvolvimento de teorias científicas e na concretização de complexos projetos tecnológicos. Foi o método de Descartes que tornou possível à NASA levar o homem à Lua. Por outro lado, a excessiva ênfase dada ao método cartesiano levou à fragmentação característica do nosso pensamento em geral e das nossas disciplinas acadêmicas, e levou à atitude generalizada de reducionismo na ciência — a crença em que todos os aspectos dos fenômenos complexos podem ser compreendidos se reduzidos às suas partes constituintes. (CAPRA, 1982, p. 54-55)

Morin e Le Moigne (2009, p. 50-52) consideram o conhecimento fragmentado um impedimento à compreensão abrangente por não permitir ligar as partes constituintes e contextualizar melhor o problema abordado. Ele chama o envolvimento do todo com suas partes de princípio hologramático da complexidade: aspectos do todo estão presentes nas partes e vice-versa. Por exemplo, a capacidade reprodutiva é encontrada tanto no ser humano (todo) quanto em suas células (parte).

Um exemplo da dificuldade para aplicar o método analítico a problemas complexos é o relato Niels Bohr (2008) ao lidar com fenômenos da física atômica:

A própria essência da explicação científica consiste na decomposição de fenômenos complexos em fenômenos mais simples. No momento, essa é a limitação essencial de que padece a descrição mecânica dos fenômenos naturais revelados pelo recente desenvolvimento da teoria atômica, que trouxe um novo interesse para esse antigo problema.  
[...] a individualidade dos efeitos quânticos típicos encontra expressão apropriada no fato de que qualquer tentativa de subdividir os fenômenos exige uma mudança do arranjo experimental, introduzindo novas possibilidades de interação entre os objetos e os instrumentos de medida, as quais, em princípio, não podem ser controladas. Consequentemente, os dados obtidos em diferentes condições experimentais não podem ser compreendidos dentro de um quadro único, mas devem ser considerados complementares, no sentido de que só a totalidade dos fenômenos esgota as informações possíveis sobre os objetos.” (BOHR, 2008, p. 5, 51)

Ao abordar o exemplo da complexidade do corpo humano, Capra (1982, p. 35; 1995, p. 137) destaca a importância de se compreender de forma holística as inter-relações e interdependências que envolvem tanto os aspectos intrínsecos do objeto de estudo quanto os aspectos referentes à interação com o ambiente natural e social. Pois não basta apenas a visão sistêmica compartimentalizada dos elementos constituintes, como numa máquina. É preciso compreender a realidade como um todo envolvente em sua máxima abrangência, o que é fundamental quando se trata de problemas complexos.

Capra (1982, p. 40) demonstra a necessidade de abordagem holística ao citar os sistemas vivos, que são organizados de tal modo que formam estruturas de múltiplos níveis, cada nível dividido em subsistemas, sendo cada um deles um "todo" em relação a suas partes, e uma "parte" relativamente a "todos" maiores. E o "todo" é considerado no sentido de serem estruturas integradas, e também "partes" de "todos" maiores, em diferentes níveis de complexidade.

No caso da física das partículas elementares, Anderson lembra que o comportamento de agregados grandes e complexos dessas partículas não é entendida em termos de uma simples extrapolação das propriedades de algumas partículas, pois “em cada nível de complexidade propriedades totalmente novas aparecem, e a compreensão dos novos comportamentos requer pesquisa que eu acho que é fundamental em sua natureza como qualquer outro.” (ANDERSON, 1972, p. 393)

Por outro lado, apesar da visão holística ser fundamental para a compreensão dos fenômenos complexos, Morin afirma que “o todo não é tudo” (2008, 159-161; MORIN, 2011, p. 75), pois a complexidade envolve também o fenômeno da emergência, onde o todo torna-se diferente das suas partes constituintes, não sendo apenas uma mera associação de elementos. E cita o caso da água - uma substância cujas propriedades são diferentes daquelas encontradas em cada uma de suas partes atômicas elementares, hidrogênio e oxigênio. (MORIN; LE MOIGNE, 2009, p. 43-44)



Um objeto integrado como um sistema organizado é relativamente autônomo dos seus constituintes individuais e sua identidade não se reduz à sua constituição. O todo se constitui além das estruturas de inter-relações e de interações de seus elementos constituintes (SANTOS, 2010, p. 20)

Segundo Fromm (2004, p. 19), a palavra 'emergência' vem do latim 'emergere', a partir da junção de *ex* + *mergere*, onde “ex” é um prefixo de origem grega, “ec”, que denota algo de fora. Portanto emergência significa algo que é lançado para fora, que se torna aparente, que vem para a luz, que se transforma e se apresenta como resultado de algum processo. Emergência é produto de alguma fusão, combinação ou junção de partes distintas ou separadas.

Um padrão emergente é uma coleção ou classe de registros que compartilham algo em comum. (FRAWLEY et al., 1992, p. 65)

Segundo O'Connor (2012), entidades emergentes (propriedades ou substâncias) “surgem” de entidades mais fundamentais, constituindo algo novo e irreduzível.

Goldstein (1999, p. 50) elenca cinco características de um padrão emergente presente num fenômeno complexo:

- 1) Novidade radical: padrões emergentes possuem características que não são previamente reconhecidas no sistema complexo sob observação. Esta novidade é a fonte da afirmação de que os recursos de emergentes não são previsíveis nem dedutíveis a partir de componentes menores ou micro-nível;
- 2) Coerência ou correlação: padrões emergentes aparecem como um todo integrado que tendem a manter algum senso de identidade ao longo do tempo;
- 3) Nível global ou macro: o *locus* dos fenômenos emergentes ocorre em nível global ou macro, em contraste com o *locus* de nível micro dos seus componentes;
- 4) Dinâmica: fenômenos emergentes não são pré-determinados em sua totalidade, mas surgem como um sistema complexo evoluindo ao longo do tempo;

- 5) Ostensivo: padrões emergentes são reconhecidos por se mostrarem espontaneamente e de forma particular, ou seja, dada a sua natureza complexa a emergência pode apresentar diferentes versões em momentos distintos.

No processo de geração de conhecimento, inferências podem auxiliar na identificação de padrões emergentes de informação (EP, *emergent patterns*) a partir de dados integrados.

## 2.2 Dado, informação e conhecimento

Dado, informação e conhecimento são termos que ocorrem juntos com frequência em textos acadêmicos<sup>3</sup>, denotando uma relação estreita entre eles, apesar das suas definições não serem muito claras em virtude da variedade de significados que lhes são atribuídos.

Há literaturas que relacionam os três conceitos como fatores essenciais e determinantes da sequência lógica de geração de conhecimento: dado gera informação, e esta, gera conhecimento. Mas esta é uma ideia simplista que não leva em consideração elementos que participam de forma concomitante desse processo como, por exemplo, contexto, semântica, capacidade do receptor, mediação pela linguagem, dentre outros.

### 2.2.1 Dado

Dentre os vários significados, abordagens e perspectivas, a Definição Geral de Informação apresentada por Floridi (2011) trata dados e informações como entidades reificadas, uma definição semelhante à de Buckland (1991), que considera informação como “coisa”.

Nesta visão, informação se baseia em dados, o que permite estabelecer uma relação mais coerente com outros termos como “mineração de dados”, “gestão da

---

<sup>3</sup> Segundo Dretske (1999, p. 45), há um núcleo comum envolvendo o significado desses termos, juntamente com outros como “inteligência”, “novidade” e “instrução”.

informação”, “base de conhecimento”, “representação de conhecimento”, gestão de documentos, e conceito formal de ontologia.

Dados são elementos essenciais discretos, que se podem distinguir de uma informação. Tal distinção é importante porque permite caracterizar a informação em elementos menos complexos que podem ser manipulados, controlados, reproduzidos e difundidos afim de gerar novos dados ou informações. Neste sentido os dados são, de acordo com Cunha e Cavalcanti (2008, p. 112-113), “a menor representação convencional e fundamental de uma informação (fato, noção, objeto, nome próprio, número, estatística, etc.) sob forma analógica ou digital passível de ser submetida a processamento manual ou automático.” E que possuem as seguintes características:

- a) podem ser estruturados de uma maneira formal;
- b) são quantificáveis;
- c) constituem sinais e códigos utilizados para alimentação, armazenamento, processamento e produção de informação;
- d) são grupos de caracteres alfabéticos, numéricos, alfanuméricos ou quaisquer outros, que representam uma condição ou valor específico.

Dado é um termo oriundo da palavra latina *datum* que significa “algo que é oferecido”, tal como se destaca na obra de Euclides, *Dedomena*, traduzida do grego para o latim.

Segundo Floridi (2011), os dados podem ser de cinco tipos:

- a) dados primários, que representam de forma direta aquilo a que se refere;
- b) dados secundários, que representa a ausência de dados, mas que produz informação;
- c) metadados: dados que atribuem significado a outros dados;
- d) dados operacionais, que atuam como controle subjacente de um sistema que produz e utiliza informação;
- e) dados derivados, que, pelo princípio da emergência da teoria da complexidade, pode ser entendido como os dados que surgem a partir de outros dados, de suas relações e combinações.

Segundo Price e Shanks (2004, p. 661), existe uma correspondência entre o conceito semiótico de “signo” e o conceito informacional de “dado”, pois ambos possuem os mesmos três componentes descritos anteriormente envolvendo sinal: uma representação de algo armazenado; a representação externa de um fenômeno como referente; e uma capacidade de interpretação pelo ser humano ou máquina. E tal interpretação real da representação dependerá das condições do intérprete, quer seja um ser humano ou uma máquina.

Segundo Santaella (2002, p. 8), uma entidade física, objeto ou fenômeno é tudo aquilo que se apresenta à percepção e à mente. E entre o objeto e a mente que interpreta existe a representação mediadora exercida pelo signo<sup>4</sup>. Por sua vez, um símbolo é um tipo de signo que designa uma realidade abstrata, que pode assumir uma forma gráfica. Por exemplo, a cruz significa o cristianismo. Porém, o símbolo ou outros signos só fazem sentido se houver capacidade de interpretação pelo interpretante. Santaella explica que, “sem regras interpretativas não há como um símbolo ter utilidade, pois este está conectado ao seu objeto em virtude de uma ideia da mente que usa o símbolo, sem o que uma tal conexão não existiria.” (p. 25)

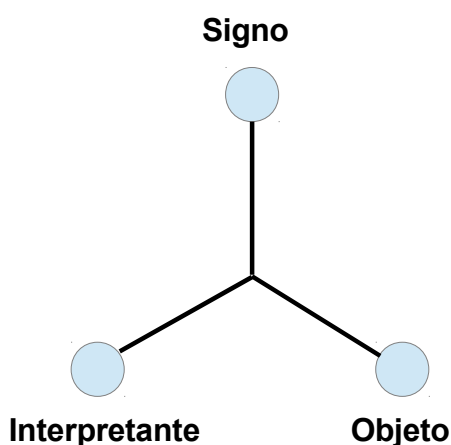


Figura 1 – Relação triádica entre interpretante, signo e objeto.

---

4 Uma entidade física serve como “signo” - um elemento da relação triádica da representação: um signo “x” representa alguma coisa “y” para algum agente “z”. (SOWA, 2000a, p. 81)

Por isso, o dado (ou conjunto de dados) que representa determinado fenômeno precisa ter capacidade informativa para poder produzir conhecimento. Ou seja, ser capaz de modificar as estruturas cognitivas (BELKIN; ROBERTSON, 1976, p. 198).

### 2.2.2 Informação

Informação é uma palavra de raízes latinas (*in formatio*), que denota o ato de dar forma a alguma coisa, que, no processo cognitivo, significa o ato de moldar a mente ou de comunicar conhecimento (CAPURRO; HJORLAND, 2007).

Segundo Bateson, informação é algo que traz novidade, que faz alguma diferença para o processo cognitivo:

Considere um homem derrubando uma árvore com um machado. Cada curso do machado é modificado ou corrigido, de acordo com a forma do corte deixado na árvore pelo golpe anterior. Este processo mental auto-corretivo é provocado pelo sistema árvore-olhos-cérebro-músculos-machado-golpe-árvore, e é este sistema total que tem as características da mente imanente.

Mais corretamente, devemos descrever este processo como: (diferenças de árvore) - (diferenças nos olhos) - (diferenças no cérebro) - (diferenças nos músculos) - (diferenças no movimento de machado) - (diferenças de árvore), etc. [...] E, como notado acima, uma diferença que faz a diferença é uma ideia ou unidade de informação. (BATESON, 1987, p. 230)

É importante ressaltar que existem processos que envolvem a formatação, o ato de dar forma pela informação, mas que não resultam na geração de conhecimento. Segundo Capurro e Hjørland (2007), a “informação não requer um processo interpretativo, embora seja uma condição necessária à aquisição de conhecimento”.

Por exemplo, no processo de formação e constituição dos seres vivos, o DNA exerce o papel de informação codificada capacitada a formatar as estruturas necessárias à vida, dentro de condições que sejam satisfatórias. Porém neste caso não há nenhuma cognição, nenhuma geração de conhecimento, tal qual ocorre na mente humana.

A relação dado-informação pode ser observada como um fenômeno complexo: a informação emerge em nível qualitativo superior a partir do conjunto dos

dados componentes, ou da falta destes, pois, conforme Floridi (2011), a ausência de dados também pode ser informação.

Além disso, segundo Sowa, o sentido do conhecimento não está nos dados armazenados, registrados, mas nas relações e combinações destes:

[...] as conexões primárias não estão nos bits e bytes que codificam os sinais, mas nas mentes das pessoas que os interpretam. O objetivo dos metadados é fazer várias conexões mentais explícitas marcando os dados com mais sinais. Esses sinais são interligações complementares que, por sua vez, também podem ser marcados por outros metadados. [...] A fonte última do sentido é o mundo físico e os agentes que usam sinais para representar entidades do mundo e suas intenções a respeito deles. (SOWA, 2000b, p. 55)

Por meio de entrelaçamentos e interligações de diferentes segmentos do conhecimento, um cientista pode criar um novo segmento de percepção e informação (FROMM, 2004, p. 28). E quando muitos dados são integrados, há grandes possibilidades de novas informações emergirem desta integração.<sup>5</sup>

Padrões emergentes de informações podem ser recuperados por meio de sistemas de descoberta de conhecimento baseados em bancos de dados (KDD - *Knowledge Discovery in Databases*), sendo fundamentais para geração de novos conhecimentos e como base para um ambiente informacional de inovação.

Segundo Fayyad et al. (1996, p.37-39), o cerne do processo de descoberta de conhecimento está na aplicação de métodos de mineração de dados (*Data Mining*) para a descoberta de padrões e de extração de informações a partir de dados mais elementares.

É importante que os dados representem de forma adequada e mais completa possível os objetos e fenômenos a que se referem, afim de que novas informações e padrões emergentes se manifestem de maneira satisfatória em meio às diversas circunstâncias e contextos situacionais (como as dimensões de tempo e espaço), e

---

5 Um exemplo histórico de emergência de novas informações a partir de integração de dados é o caso da descoberta do principal foco de cólera na cidade de Londres em 1854, relatado por Johnson (2008).

Naquela época não se conhecia bem o que provocava o cólera, nem as formas e os lugares de contaminação. Tratava-se de um agente patogênico letal que se tinha poucas informações a seu respeito. Como resultado, houveram grandes epidemias da doença, particularmente nas cidades. Neste tempo, dois cientistas recolheram dados a partir de ocorrências da doença e os integraram na composição de um mapa. Deste mapa “emergiu” informações importantes até então não identificadas que culminaram na localização do principal foco da doença: uma fonte de água altamente contaminada e muito utilizada pelos cidadãos londrinos.

às necessidades e capacidades interpretantes de seus usuários.

Pierre Levy lembra que “quando mensagens fora de contexto e ambíguas começam a circular, a atribuição do sentido passa a ocupar um lugar central no processo de comunicação.” (1997, p. 89)

Segundo Nonaka et al. (2000, p. 7), por ser dinâmico, o conhecimento é específico do contexto, uma vez que depende de um determinado tempo e espaço. E informação sem contexto é apenas informação, não conhecimento.

Alguns autores defendem a ideia de que sem representação de dados não há informação (FLORIDI, 2011). A interpretação da representação dada pela informação é algo relativo: a mesma representação de um objeto pode conter diferentes informações para diferentes pessoas:

informação é o que é informativo para uma determinada pessoa. O que é informativo depende das necessidades interpretativas e habilidades do indivíduo (embora estas sejam frequentemente compartilhadas com membros de uma mesma comunidade de discurso)

[...] a informação é um conceito subjetivo, mas não fundamentalmente em um sentido individual. Os critérios sobre o que conta como informação são formulados por processos socioculturais e científicos. Usuários deveriam ser vistos como indivíduos em situações concretas dentro de organizações sociais e domínios de conhecimento. Uma pedra em um campo pode conter diferentes informações para pessoas diferentes. Não é possível para os sistemas de informação mapear todos os possíveis valores de informação. (CAPURRO; HJORLAND, 2007, p. 155, 192)

Para fazer sentido, o conhecimento registrado precisa ser descrito e contextualizado por meio de outras informações subjacentes de forma que represente da melhor forma possível o conhecimento tácito que o gerou. (TUOMI, 1999, p. 11)

### **2.2.3 Conhecimento**

Informação e capacidade de interpretação são elementos fundamentais para o processo cognitivo. Do ponto de vista do processo mental (interno) de geração de conhecimento, é de natureza externa. Pode ser tangível, manifestando-se como uma expressão, descrição ou representação do conhecimento com a finalidade de compartilhamento, fundamental para a geração e crescimento de conhecimento. Neste sentido, afirma Popper:

Toda a experiência consiste de informação recebida através de nossos sentidos... o conhecimento está, antes de tudo, em nós: informação que nos atingiu e que conseguimos absorver ... O crescimento de todo conhecimento consiste na modificação de conhecimento prévio. O conhecimento nunca começa do nada, mas sempre de algum conhecimento de base. (POPPER, 1999, p. 67, 71, 75)

Polanyi (1993) classificou o conhecimento humano em duas categorias: conhecimento “explícito” ou codificável, que pode ser transmitido numa linguagem formal e sistemática; e o conhecimento tácito, intrinsecamente pessoal, difícil de formalizar ou comunicar (NONAKA, 1994, p. 16; MCNERNEY, 2002, p. 1011).

Takeuchi e Nonaka afirma que conhecimento tácito e explícito são a mesma coisa e que a distinção é apenas aparente (2008, p. 20). O que existe é um processo cíclico de transição de conhecimento tácito para explícito e vice-versa, chamado de “criação e dialética do conhecimento”.

O conhecimento (tácito) é criado a partir da informação (conhecimento explícito) e, ao socializar-se transforma-se em novamente em informação (BARITÉ, 2001).

Segundo Nonaka e Takeuchi (1995, p. 61-73), a conversão do conhecimento tácito para conhecimento explícito e vice-versa pode ocorrer de quatro modos de conversão:

- 1) Socialização - de conhecimento tácito para conhecimento tácito.  
Exemplo: conversas entre pessoas;
- 2) Externalização - de conhecimento tácito para conhecimento explícito.  
Por exemplo registro de dados falados ou gerados por uma fonte;
- 3) Internalização - de conhecimento explícito para conhecimento tácito.  
Por exemplo: leitura de dados registrados;
- 4) Combinação - de conhecimento explícito para conhecimento explícito.  
Por exemplo: agregação e classificação de dados registrados para produção de novos dados.





Figura 2 - Quatro modos de conversão do conhecimento (adaptado de NONAKA; TAKEUCHI, 1995, p. 62, 71).

### 2.2.3.1 Conhecimento tácito

O conhecimento tácito é aquele que é processado na mente e que compreende as experiências, percepções e informações adquiridas. É difícil de ser expressado em sua totalidade e que, por isso, “podemos pensar muito mais do que podemos dizer” (POLANYI, 1983, p. 4). Ou, nas palavras de Brabandere, “o que pode ser pensando tão profundamente é impossível de expressar” (2006, p. 38). Ou ainda, o conhecimento é informação intangível, que não pode ser tocada, nem medida de uma forma direta. É pessoal, subjetiva e conceitual (BUCKLAND, 1991, p. 1).

Polanyi (op. cit., p. 5-14) e Takeuchi e Nonaka (2008, p. 19) afirmam que uma das características do conhecimento tácito é de ser de difícil comunicação e compartilhamento, pois o conhecimento tácito engloba as experiências, os valores, as emoções, as habilidades, modelos mentais, percepções, ou outros conhecimentos incorporados no indivíduo. Para Choo (2006, p. 135), o conhecimento tácito é difícil de ser verbalizado porque “é expresso através de uma

ação com base em habilidades e não pode ser reduzido a regras e receitas.”

### 2.2.3.2 Conhecimento explícito

Segundo Cunha e Cavalcanti (2008, p. 102; 132), o conhecimento explícito é o conjunto de informações registradas e estruturadas para a compreensão humana baseadas em algum suporte. Definição esta que é válida tanto para os documentos em papel (substanciais) quanto os documentos eletrônicos (insubstanciais).

Para LANKHORST et al. (2005, p. 72) “o conhecimento explícito refere-se ao conhecimento que pode ser exteriorizado em termos de alguma representação. Com a representação do conhecimento, nos referimos ao processo de codificação de conhecimento em termos de uma linguagem em algum meio, por exemplo, a criação de um modelo de arquitetura.”

Buckland (1991, p. 353-354) considera “dados” como os registros armazenados em em computador como textos, documentos e objetos, informação tangível, ou seja, a informação como “coisa”, algo que, apesar de imaterial, possui características bem definidas.

Porém, existem dois problemas envolvendo o conhecimento explícito em relação ao tácito:

- a) a dificuldade de transformar conhecimento tácito em explícito, o que faz deste último um subconjunto do conhecimento tácito. Pois a representação é um processo limitante. Informação reificada (como coisa) corresponde à apenas uma parte da totalidade de significados do objeto ou fenômeno que representam. Buckland lembra que o registro de uma informação é uma representação limitada do objeto ou fenômeno:

Cada representação pode ser esperada a ser mais ou menos incompleta em algum respeito. Uma fotografia não indica movimento e podem não representar a cor. Mesmo uma fotografia colorida, em geral mostram cores imperfeitamente - e desaparecer com o tempo. Uma narrativa escrita refletirá o ponto de vista do escritor e as limitações da língua. Filmes e fotografias mostram geralmente só uma perspectiva. Algo do original é sempre perdido. Há sempre alguma distorção, mesmo que apenas através de incompletude (BUCKLAND, 1991, p. 358).

- b) o aspecto estático, caracterizado pela rigidez e fixação do conhecimento explícito, restringe a flexibilidade de seu uso no processo dinâmico de

geração de novos conhecimentos e inovação, em "um mundo que não está pronto, mas em construção" (TUOMI, 1999, p. 16);

Este problemas de transição entre conhecimento tácito e explícito são particularmente importantes no âmbito das organizações corporativas, uma vez que tanto o conhecimento das pessoas quanto o conhecimento cristalizado em algum suporte são fundamentais para que as ações corporativas sejam executadas da melhor maneira possível. Por isso, as organizações procuram maximizar a recuperação, armazenamento e uso do conhecimento (WIIG, 1997, p. 8).

### **2.3 Conhecimento organizacional e gestão do conhecimento**

Uma organização é uma entidade social que realiza atividades de esforço comum, combinado e coordenado, em sinergia, para alcançar objetivos determinados.

O conhecimento organizacional pode estar distribuído entre diversas fontes sob diversos pontos de vista (DIENG-KUNTZ; MATTA, 2002, p. xvi).

O conhecimento organizacional está disposto em duas dimensões: uma dimensão estática (sua estrutura); e uma dimensão dinâmica (processos e atividades).

O lado dinâmico compreende os processos que geram, armazenam, combinam e distribuem o conhecimento e, pelo lado estático, a estrutura das fontes do conhecimento, suas características e suas relações mútuas (SPEK; SPIJKERVET, 2005, p. 18-22).

Segundo Spek e Spijkervet (op. cit.), os processos do conhecimento organizacional podem ser compreendidos como quatro operações básicas (Figura 3):

- a) combinação do conhecimento disponível; o melhor da organização ocorre quando o conhecimento disponível, multidisciplinar, pode ser combinado;
- b) desenvolvimento de novos conhecimentos: surge a partir de ideias

criativas, aprendizado com os erros, experiência cotidiana e o esforço dos departamentos de pesquisa e desenvolvimento;

- c) proteção do conhecimento para a organização: o conhecimento individual deve se tornar acessível a toda organização, sendo disponibilizado da forma correta, no momento certo, de acordo com as necessidades;
- d) distribuição de conhecimentos: o conhecimento deve ser distribuído o mais rápido possível para aqueles que dele necessitam.

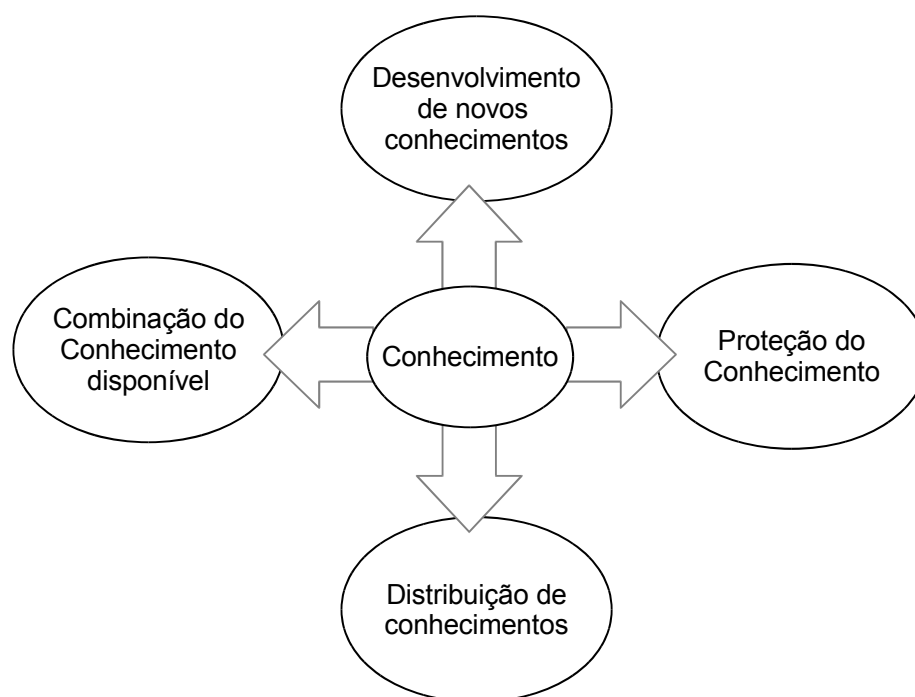


Figura 3 - As quatro operações básicas do conhecimento organizacional, segundo Spek e Spijkervet (2005, p. 18)

A estrutura do conhecimento organizacional compreende características:

- a) forma: denota quais são as fontes de conhecimento. Neste caso, as pessoas são fontes de conhecimento (tácito). São agentes ativos capazes de aplicar, desenvolver e aumentar o conhecimento por meio das suas atividades. A outra forma de conhecimento é o explícito, armazenado em fontes passivas como os livros, documentos e outras formas de registro e suporte;

- b) local: o conhecimento pode estar localizado dentro ou fora da organização, ou distribuído ou pulverizado pela organização;
- c) tempo: indica o período de uso do conhecimento. Por exemplo, em caso de necessidade, computadores podem ser utilizados para gerar rapidamente conhecimento a partir da análise de grande volumes de informação;
- d) conteúdo: a estrutura de conteúdo denota protocolos, regras e procedimentos que atam sobre o conhecimento organizacional.

Afim de garantir ações inteligentes no contexto das duas dimensões (processos e estruturas), as organizações desenvolvem processos sistemáticos, articulados, que contribuem para efetiva identificação, geração, compartilhamento e aplicação do conhecimento, aos quais, em seu conjunto, dar-se o nome de gestão do conhecimento (CUNHA; CAVALCANTI, 2008, p. 180).

De acordo com Steels (1993, apud DIENG et al., 1998), o principal objetivo de gestão do conhecimento (KM) em uma organização é promover o crescimento, a comunicação e a preservação do conhecimento na organização.

Numa perspectiva estratégica, Spek e Spijkervet (*op. cit.*, p. 25) afirmam que a gestão do Conhecimento concentra-se nas seguintes tarefas:

- a) formulação de uma política estratégica para o desenvolvimento e aplicação do conhecimento;
- b) implementação de uma política do conhecimento com o apoio de todas as partes da organização;
- c) melhorar a organização onde o conhecimento não é utilizado de forma otimizada, ou não está a ser adaptado para novas circunstâncias;
- d) monitorar e avaliar as realizações de ativos de conhecimento e atividades de gestão em termos de conhecimento.

O'Leary (1998) ressalta que a gestão de conhecimento procura facilitar o acesso e o reuso do conhecimento organizacional, normalmente por meio de tecnologia de informação avançada.

A gestão do conhecimento procura resolver os problemas das transições

entre conhecimento tácito com explícito e vice-versa. Particularmente, a combinação de conhecimento explícito com outro conhecimento explícito pode levar a novos conhecimentos, mediante a reconfiguração das informações existentes através da classificação, acréscimo, recategorização e recontextualização do conhecimento explícito (NONAKA, 1994, p. 19).

Neste contexto de combinação entre conhecimentos explícitos, é fundamental o desenvolvimento de uma memória organizacional, “um dispositivo que permite o registro, a conservação e a restituição de dados. No âmbito das organizações, a memória é considerada como um repositório institucional de conhecimento” (CUNHA; CAVALCANTI, 2008, p. 243-244).

## **2.4 A memória das organizações**

A memória organizacional pode ser caracterizada como o conjunto das experiências históricas codificadas na estrutura organizacional como documentos, bases de dados e outras formas de armazenagem.

Segundo Moser et al., (2009, p. 123) muitas vezes a memória é o único fundamento de que dispomos para afirmar que conhecemos alguma coisa. É a memória que está por trás de boa parte dos conhecimentos que supomos ter, conhecimentos que de outro modo não teriam fundamento algum. E a perda de memória pode levar a uma perda de conhecimento.

De acordo com Tuomi (1995, p. 303), a memória institucional constitui os entendimentos e crenças armazenados e compartilhados pelos indivíduos de uma organização, sendo um requisito fundamental para a comunicação do conhecimento. A informação registrada é impessoal, passível de ser compartilhada, como documentos e bancos de dados, podem ser considerados como memória institucional. Ainda, segundo este autor, a memória organizacional é fruto da necessidade de lidar com o aumento da complexidade do ambiente social e suas interações:

à medida que a complexidade do ambiente e suas mudanças sobrecarreguem a capacidade de aprendizagem individual, a aprendizagem organizacional se torna mais dependente da percepção, comunicação, e interpretação coletivas, e menos dependente de memórias particulares que residem nas cabeças dos seus membros. (TUOMI, 1995, p. 305)

Segundo Dieng-Kuntz et al., a memória corporativa é definida como:

a explícita e persistente materialização da informação e do conhecimento cruciais para uma organização, afim de facilitar o seu acesso, compartilhamento e reutilização pelos membros da organização em suas tarefas individuais e coletivas (DIENG-KUNTZ et al., 2001 apud DIENG, 2005, p. 67).

Abecker et al. (1998, p. 41-42) consideram que a memória organizacional promove a acumulação, estruturação e utilização do conhecimento explícito em múltiplas formas por meio da captura e organização das informações oriundas de várias fontes dados. Uma memória institucional provê informação útil que deve ser disponibilizada prontamente aos usuários, sendo capaz de se adaptar às necessidades informacionais. O objetivo principal é evitar perdas e aumentar a acessibilidade de todos os tipos de conhecimento em nível corporativo.

Destaca-se, porém, o fato de que a preservação e gestão da memória organizacional estão entre as atividades mais negligenciadas pelas modernas organizações que, inundadas por informações, padecem de amnésia organizacional (CHOO, 2002, p. 36). E uma memória deficiente implica em dificuldades de aprendizado, julgamento, interpretação, além de problemas de se contextualizar eventos (EYSENK; KEANE, 1990 apud TUOMI, 1995, p. 304).

Nesse contexto, Donn Neal afirma:

Enquanto as corporações estão sendo sobrecarregadas por um volume crescente de registros, são, paradoxalmente, incapazes muitas vezes de localizar a informação verdadeiramente vital. Sem um adequado programa de arquivamento, registros importantes podem ser perdidos, sem medidas de conservação adequadas, muitos documentos valiosos, são literalmente auto-destruídos. Sem precauções, um desastre imprevisto pode eliminar informações insubstituíveis. E mesmo quando os registros são mantidos, não há garantia de que sejam localizados. A falta de um programa eficaz de arquivamento pode tornar difícil para a empresa responder com rapidez e precisão não apenas às suas próprias necessidades de informação, como também para legitimar consultas externas - de órgãos governamentais, os consumidores, e a mídia. As corporações vão crescer cada vez mais incapazes de resolver estes desafios sem um eficiente e abrangente programa de arquivamento. (NEAL, 1993 apud CHOO, 2002, p. 36)

A memória institucional é um componente importante da gestão do conhecimento por possibilitar a comunicação e a reutilização de conhecimento entre os diferentes membros da organização. Neste sentido Guarino acrescenta:

A necessidade de ferramentas adequadas para a organização da informação, extração e integração tornou-se mais e mais evidente. Numa perspectiva de "informação global", o valor acrescentado de um pedaço de informação codificado não é mais apenas limitado a uma determinada aplicação que motivou sua aquisição, mas tende a aumentar na dependência de sua reutilização, ou seja, a sua adequação para ser dinamicamente integrados dentro de vários corpos diferentes de informações." (GUARINO, 1997, p. 139)

Exemplificando as palavras de Guarino, um ponto de geoprocessamento coletado, uma sequência de dados escrita numa caderneta de campo, uma série de dados coletado em laboratório sob condições adversas, todas essas informações são obtidas a certo custo e têm seus valores ampliados quando são integradas e reaproveitadas de alguma forma para outros contextos e usos.

A memória corporativa é derivada das experiências da organização ao interagir com o ambiente, sendo expressa formalmente sob a forma de documentos e procedimentos (CHOO, 2002, p. 17-18). Neste sentido, a manutenção de uma memória organizacional depende da gestão do seu conteúdo. O conhecimento é um fenômeno dinâmico e, como consequência, novos conhecimentos explícitos são gerados e registrados em crescente em volume. E, na mesma medida, também cresce a necessidade de melhor armazenamento, organização e capacidade de acesso.

## **2.5 A gestão de documentos**

A informação pode ter origem de diversas fontes. Alguns autores consideram que qualquer objeto material pode ser considerado como fonte: um animal, como um antílope (BRIET apud BUCKLAND, 1991), uma pedra, uma escultura, dentre outros, podem emitir alguma mensagem informativa, sem a necessidade de se distinguir seus elementos constituintes (os dados), pois a informação é um fenômeno



emergente.

Por outro lado, a informação pode não ser um objeto em si, mas uma representação registrada e armazenada em algum suporte físico ou eletrônico, que estabiliza seu conteúdo no tempo e no espaço.

Cunha e Cavalcanti (2008, p. 132-133) definem documento como um suporte de informação, uma representação da realidade, registrada e ou codificada sob a forma de textos, fotografias e outros tipos de imagens, filmes, sons e músicas. Um objeto com capacidade informacional que pode ser manipulado.

Para David Levy (1998, p. 153), documentos falam de coisas: “são pedaços do mundo material - argila, pedra, pele animal, fibra vegetal, areia - que temos imbuídos com a capacidade de falar. Eles são substitutos ou agentes que criamos para falar por nós ou em nosso nome.”

Documentos são formas de manter o conhecimento explícito. Ranganathan considera a informação tangível como pensamento incorporado em algum suporte, um documento, meio pelo qual o conhecimento pode ser manipulado, transportado através do espaço e preservado através do tempo (BUCKLAND, 1997, p. 807).

Um documento é informação registrada e acessar um documento é uma forma indireta de acessar informação. Neste sentido, van Rijsbergen considera que:

Um sistema de recuperação de informação não informa (isto é, muda o conhecimento sobre algo) ao usuário sobre o assunto de sua investigação. Ele meramente informa sobre a existência (ou não) e a localização dos documentos relacionados ao seu pedido (VAN RIJSBERGEN, 1979, p. 1 apud CAPURRO; HJORLAND, 2007).

O volume de documentos tem crescido em grande escala tanto em quantidade quanto de qualidade. Textos e imagens são produzidos e compartilhados com maior facilidade, imagens são geradas com maior qualidade de nitidez e cores em virtude da alta capacidade de processamento das TICs. Mas a capacidade de acesso e assimilação dos documentos não tem acompanhado esse crescimento.

Documentos textuais e imagens são fontes de informação destinadas à internalização por parte das pessoas, onde o conhecimento explícito interpretado gera conhecimento implícito. Cada texto, imagem ou som registrados em documentos constituem um todo de difícil redução aos seus elementos que

poderiam revelar algum significado referente ao fenômeno por eles representados.

O conteúdo dos documentos nem sempre apresenta elementos componentes dispostos em estruturas que possam ser combinados com outros conteúdos a fim de constituir uma nova informação em nível mais elevado, emergente.

A problemática envolvendo documentos é contextualizada por Rosner et al.:

Ao longo de seu ciclo de vida, os documentos são processados por computadores: eles são criados, modificados e atualizados com processadores de texto, são armazenados e recuperados a partir de bases de dados de documentos, são distribuídos através de redes, e podem estar sujeitos a transformações para impressão ou exibição. Do ponto de vista da gestão do conhecimento, o processamento de tais documentos, tem uma grande limitação: dificilmente capta a riqueza do conhecimento contida nesses documentos. Em particular, não há como capturar o conhecimento implicitamente referidos no documento. E mesmo o conhecimento que está explicitamente expresso é apenas uma utilidade limitada, uma vez que não pode ser processado de forma automática e, assim, ser explorado por diferentes aplicações como, por exemplo, geradores automáticos de documentos. (ROSNER et al., 1997, p. 955)

Para facilitar a manipulação e o acesso de grandes volumes de informação, principalmente por meio de máquinas, é preciso criar esquemas de organização – formatos, padrões, estruturas e informações subjacentes para facilitar o acesso aos documentos.

O processo de internalização proposto por Nonaka e Takeuchi (1995), baseado na assimilação do conhecimento explícito (em algum suporte) para o conhecimento implícito (tácito, da mente humana), pode ser compreendido a partir da relação do trinômio “Documento-Informação-Usuário”, presente no processo de um sistema de recuperação de informação que “indica como objetivo fim a satisfação do usuário na busca de conhecimento registrado em documentos.” (SOUZA, 2007, p. 113)

É complexo determinar qual informação é importante e relevante devido a muitos aspectos envolvidos. O problema é que a simples relação “Documento-Informação-Usuário” é complexa, pois implica, de um lado, uma variedade de documentos, e de outro, imprecisas e variáveis necessidades e capacidades de assimilação por parte do usuário.

Segundo Nonaka (1994, p. 17), o significado da informação difere de acordo com o contexto e do que se pretende fazer com ela.

Partindo do conceito de documento e da definição de informação como “coisa”, Buckland (1991, p. 356) afirma que, para um documento ser informativo, sua relevância depende das circunstâncias, das situações e das condições do pesquisador e da pesquisa que está sendo feita na busca por informação.

Segundo Gluck (1996, p. 91-94), a relevância é um conceito multidimensional, baseado no contexto ou situação, e na dinâmica sobre como resolver as necessidades dos usuários. Depende também das condições de recepção, do estado cognitivo do usuário, ou seja, de sua capacidade de assimilação e interpretação. Como resultado, a resposta pode ser a informação solicitada ou não, pois o mecanismo de busca e recuperação esteja funcionando de forma perfeita e precisa, não garantia de que as necessidades de informação sejam satisfeitas.

Segundo Vickery e Vickery:

apenas o destinatário final de uma mensagem de informação pode fazer um julgamento inequívoco e a informação somente é gerada **se for potencialmente relevante para as necessidades conhecidas ou hipotéticas**, que só serão registrados se um uso posterior no tempo está previsto.(VICKERY; VICKERY, 2004, p. 265, grifo nosso)

Para Hjørland, “o tipo de informação que é julgado relevante para uma determinada tarefa depende da pessoa que faz o julgamento.” (2008c, p. 18)

A relevância é um conceito ligado com os interesses, expectativas e necessidades de uma pessoa ou de um sistema ou módulo de requisição de busca automática de algum sistema de informação. Segundo Schamber et al.:

“Relevância é o juízo da qualidade da relação entre informação e necessidade de informação do usuário... é uma determinação de como os usuários percebem a informação relativa às suas situações necessidade de informação.

[...] Necessidade de informação é definida como a forma como os usuários percebem seus ambientes situacionais como sendo obscuros (descontínuos ou incertos) em conjunto com a forma como eles percebem a informação como meio de ajudá-los a esclarecer ou dar sentido a esses ambientes ... o *locus* de relevância está dentro de percepção do indivíduo de informações e informações sobre o ambiente, não em informações como estão representadas em um documento ou alguma outra forma concreta.” (SCHAMBER et al., 1990, p. 770-771)

Neste aspecto Saracevic (1970, p. 121-122) levantou três hipóteses:

- a) Somente o próprio usuário pode julgar a relevância dos documentos para ele e seus usos;

- b) Para o usuário mesmo um julgamento relevância pode mudar ao longo do tempo;
- c) Vários tipos de julgamentos podem existir por causa das diferentes propósitos para os quais a informação é necessária.

Schamber et al. (1990, p. 771) descrevem ainda a natureza dinâmica e complexa da relevância em quatro aspectos:

- a) “o mesmo item de informação significa coisas diferentes para pessoas diferentes ao mesmo tempo e coisas diferentes para o mesmo indivíduo em momentos diferentes”;
- b) “os juízos de relevância são as avaliações dos usuários de informação (a partir de fontes internas ou externas) em relação à sua situação de necessidade de informação em pontos específicos no tempo”;
- c) “Relevância, então, é um conceito dinâmico que depende julgamentos dos usuários individuais da qualidade da relação entre a informação e a necessidade de informação a um certo ponto no tempo”;
- d) “critérios de relevância estão interligados no tempo e espaço, e internamente construído na mente do usuário.”

Dadas estas características dinâmicas e complexas pode-se concluir que uma informação pode ser relevante em algum momento para determinado receptor, numa determinada situação, circunstância, ou contexto.

Um conjunto de números, uma imagem, a descrição de um evento ou de um fenômeno é passível de relevância e, portanto, deve ser registrado para ser consultado. Um objeto aparentemente sem importância pode vir a ser um componente fundamental em alguma pesquisa futura. Uma série de dados antigos podem servir de parâmetros de referência para novas pesquisas. Uma informação pode ser irrelevante de forma isolada. Porém, agregada com outras informações, podem gerar padrões informacionais que revelam um novo conhecimento emergente até então desconhecido.

Neste contexto, o volume das informações tende a aumentar, considerando-se a perspectiva de que toda e qualquer informação seja potencialmente importante

e deva ser registrada, e que os problemas de armazenamento e processamento são mitigados pela melhoria das TICs a custos cada vez menores.

O novo desafio passa a ser como organizar e representar o conhecimento científico registrado de forma que possam ser recuperado sob qualquer circunstância, atendendo precisamente as necessidades do usuário.

É preciso estabelecer alguma forma de organização para que seja possível a recuperação da informação (TAYLOR; JOUDREY, 2009, p. 2). O conhecimento sempre se registra em documentos e deve ser organizado para que seja melhor aproveitado tanto no âmbito individual quanto social. (BARITÉ, 2001)

Além disso, grandes volumes de informação que estejam disponíveis em algum suporte, necessitam de algum mecanismo automático que estabeleça a intermediação entre a fonte de informação e o usuário. A disponibilidade de numerosos livros em uma biblioteca ou o grande volume de dados de um banco de dados corporativo são exemplos de conhecimento explícito armazenado. Mas não há garantias de que este conhecimento seja efetivamente assimilado, transformando-se em conhecimento implícito. Isto dependerá da forma como foi organizado e dos mecanismos de acesso empregados.

## **2.6 Organização do conhecimento**

O termo “organização” vem do latim “*organum*” e significa “instrumento para uma ação específica” (ABBAGNANO, 2007, p. 855). Portanto, pode-se afirmar que a organização do conhecimento é uma série de instrumentos que visam facilitar o acesso da informação nos documentos por parte do usuário. Souza (2007) complementa que, entre o documento e o usuário, há um conjunto de atividades essenciais que possibilitam a busca e o acesso pela informação: classificação, indexação, recuperação e disseminação.

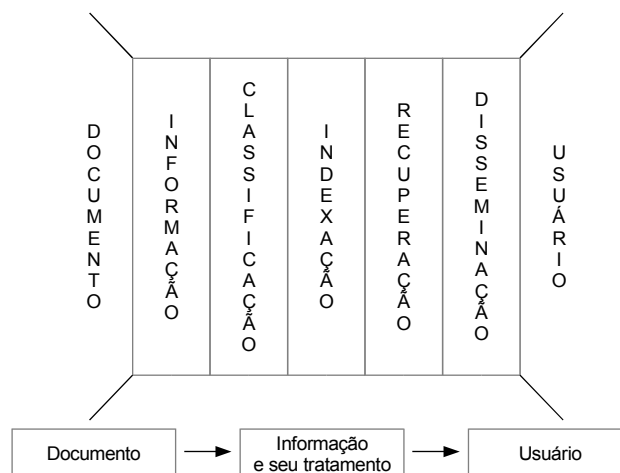


Figura 4 - Elementos de um sistema de informação (SOUZA, 2007, p. 113).

A classificação é um processo que permite agregar objetos em classes, definidas em torno de pelo menos uma propriedade em comum (CUNHA; CAVALCANTI, 2008, p. 83).

Langridge afirma que:

sem classificação não poderia haver nenhum pensamento humano, ação e organização que conhecemos. A classificação transforma impressões sensoriais, isoladas e incoerentes em objetos reconhecíveis e padrões recorríveis. (LANGRIDGE apud SOUZA, 2000)

O processo de recuperação do conhecimento contido em documentos é diferente da recuperação de dados elementares. Os textos, por exemplo com sua composição léxico-semântica, formam um todo coerente destinado à leitura pelas pessoas, sendo difícil a distinção de seus componentes por partes dos computadores. Da mesma forma, registros de sons e imagens, e objetos reais em três dimensões emanam conhecimento explícito que é pouco distinguido por máquinas.

Uma forma de permitir a acessibilidade dos conteúdos de tais documentos é utilizar informações subjacentes que representem o conhecimento ali contido. Estas informações subjacentes recebem o nome de “metadados”, ou seja, dados sobre os dados.

Um metadado atribui significado a uma informação permitindo seu acesso e

sua compreensão. A descrição de informações por meio de metadados, e os processos de classificação e indexação, são formas de organização do conhecimento, que proporcionam aos usuários melhor capacidade para encontrar informações relevantes.

Haynes (2004 apud ROWLEY; HARTLEY, 2008, p. 43-44) elenca cinco propósitos para os metadados: descrição do recurso; recuperação da informação; gestão do ciclo de vida do documento (autoria, datas de criação, modificação e publicação, etc.); definição de propriedade e autenticidade; e interoperabilidade (conexões entre documentos, informações e fontes de informações).

Segundo Sicilia e Lytras (2009, prefácio, p. v), existem pelo menos três níveis de crescente funcionalidade dos metadados:

- a) dar suporte à descoberta de informações;
- b) facilitar a compreensão de aquisição, e a utilização das informações por seres humanos; e
- c) permitir que as informações possam ser descobertas, gerenciadas, processadas e analisadas de forma automática.

A indexação é uma forma de facilitar o acesso aos documento por meio da descrição sucinta e pela definição de termos referentes ao seu conteúdo. A descrição por meio de metadados agrega significado (semântica) ao objeto referenciado (documento, arquivo, conjunto de dados, dado elementar) de forma a caracterizar e externalizar aspectos relevantes que facilitem a identificação e o acesso.

De acordo com Sowa, semântica é o estudo da relação entre as indicações para as coisas do mundo e os padrões de sinais correspondentes que ocorrem entre as coisas às quais se referem. (2000b, p. 56)

Segundo Cunha e Cavalcanti (2008, p. 330), semântica é o “estudo da relação de significação nos signos e da representação do sentido dos enunciados”, ou o “estudo das relações entre os signos e os objetos por eles designados”.

A semântica de dados constitui toda definição e caracterização sobre o que os dados significam a partir das relações entres eles ou a partir de dados que os

descrevem (metadados).

Porém, dada a subjetividade do processo de organização, Hjørland e Pedersen (apud HJORLAND 2008b) lembram algumas características importantes quanto a classificação e descrição, das quais destacam-se:

- a) a classificação estabelece uma ordem às coisas por meio da definição de classes, com base em algumas propriedades;
- b) as propriedades das coisas não estão totalmente disponíveis (dadas), mas apresentadas sob a forma de algumas descrições e pré-entendimentos;
- c) a descrição (ou qualquer outro tipo de representação) é subjetiva, pois é um reflexo tanto da coisa descrita quanto da percepção do sujeito que a descreve;
- d) a seleção das propriedades das coisas reflete a finalidade da classificação;
- e) é a falsa crença de que os critérios para a classificação sejam objetivos;
- f) diferentes domínios de conhecimento podem necessitar de diferentes descrições e classificações;
- g) em todos os domínios existem diferentes teorias, abordagens, interesses, ou “paradigmas”, que tendem a descrever e classificar as coisas de acordo com suas respectivas visões e objetivos.

Uma forma de se reduzir tanto a complexidade do conhecimento quanto a subjetividade do processo de organização é estabelecer mecanismos mais formais e consensuais de organização, por meio de vocabulário controlado - um conjunto estruturado de termos e conceitos bem definidos de forma consensual, não ambíguos ou arbitrários, para uso compartilhado – que possibilita a uniformização, formalização e padronização dos elementos informacionais que fazem referência ao conhecimento que se deseja organizar.

Segundo Allemang:

A ideia de um vocabulário controlado é que um conjunto de termos serão selecionados e geridos por um indivíduo ou grupo na organização social. O vocabulário controlado representa uma política de toda a organização sobre quais termos serão utilizados para se referir a certas coisas, e sua normalização (por exemplo, ortografia, numeração, etc). (ALLEMANG, 2010, p. 8)



No âmbito de uma organização social, sem um vocabulário controlado uma mesma palavra (termo) pode ter diferentes conceitos e diferentes palavras podem se referir ao mesmo conceito. Sem um vocabulário compartilhado, a informação distribuída provavelmente permanecerá isolada e aplicações computacionais terão dificuldades para integrar informações e processar inferências a partir de termos e relações semânticas, que poderiam ser capazes de gerar novas informações e conhecimentos.

Kobashi complementa:

A localização ou identificação de informação, sem padronização léxica, torna-se errática. Resultados eficientes de busca dependem, assim, de coincidência entre as formas de representação utilizadas pelo sistema de informação e pelo usuário. Um vocabulário controlado, portanto, garante a comunicação efetiva entre sistema de informação e usuário. (KOBASHI, 2008, p. 1)

Segundo Hjørland (2007a), “a ideia de um vocabulário controlado é a de reduzir a variabilidade de expressão utilizada para caracterizar o documento a ser indexado, evitando sinônimos e removendo a ambiguidade (homônimos).”

Zeng classifica quatro importantes princípios que regem os vocabulários controlados:

- a) eliminação da ambiguidade, que ocorre quando uma palavra ou frase possui mais de um significado. No vocabulário cada termo possui apenas um significado;
- b) controle de sinônimos, onde um termo principal é definido e associado a outros termos, com todos referenciando um mesmo conceito;
- c) estabelecimento de relacionamento entre termos, com a definição de relações semânticas (de significado), tais como equivalência, hierarquia e associação;
- d) possibilidade de teste e validação dos termos, onde estes são definidos e selecionados por meio de critérios, e sua composição pode ser revisada.

O uso de vocabulário controlado traz três benefícios diretos para o compartilhamento do conhecimento:

- a) integração entre diversas informações ou fontes referenciadas pelo

vocabulário;

- b) interoperabilidade semântica: padrões de descrição estabelecem uma concordância entre os conteúdos informacionais (ROWLEY; HARTLEY, 2008, p. 43-44);
- c) capacidade de processamento de inferências, fundamental em bases de conhecimento, ou seja, em bases de dados que permitem a combinação de seus conteúdos, gerando novos conhecimentos emergentes ou a localização indireta (não-trivial) de informações correlatas e subjacentes.

Por suas características, os vocabulários controlados são utilizados em esquemas que “apresentam a interpretação organizada das estruturas de conhecimento”, chamados de Sistemas de Organização do Conhecimento (KOS, inglês, *Knowledge Organization Systems*) (HJORLAND, 2008a). Dentre estes esquemas destacam-se taxonomias, tesouros e ontologias (WOOD, 2010, p. 285).

Na organização do conhecimento os KOS facilitam a descrição, integração, acesso e processamento do conhecimento explícito. Em síntese, enquanto a organização do conhecimento é um processo que auxilia o acesso ao conhecimento, os KOS são as ferramentas empregadas para sua aplicação.

De acordo com Hodge (2000), há algumas características comuns entre os KOS:

- a) os KOS são representações, ou seja uma visão particular do mundo;
- b) a mesma entidade pode ser caracterizada de formas diferentes dependendo do tipo de KOS utilizado; e
- c) deve haver semelhança suficiente entre o conceito expresso pelo KOS e o objeto do mundo real (referente).

Hodge (2000; HJORLAND, 2008b) classifica os tipos de sistemas de organização do conhecimento da seguinte forma:

- a) listas de termos: lista autorizada, glossários, dicionários, *gazetteers* (lista de lugares);
- b) classificações e categorias: cabeçalho de assunto; esquemas de

classificação, taxonomias; esquemas de categorização;  
c) listas relacionais: tesouro, redes semânticas, ontologias.

O conhecimento explícito, bem como os sistemas de organização de conhecimento subjacentes, possui natureza estática quando comparado com o conhecimento implícito, complexo e de natureza dinâmica, que está sempre mudando. Como afirma McInerney (2002): “ao contrário da informação estática, que pode ser realizada em bases de dados e em papel, o conhecimento é baseado em seres conscientes, ou emana deles, e assim, está sempre mudando com a experiência humana.”

O conhecimento implícito é fruto de processos cognitivos tanto individual (interno, na mente), quanto de processos sociais (interativos) influenciados pelo ambiente, contextos, capacidades de assimilação e interpretação. Porém, o conhecimento explícito é uma representação estática, apesar da sua importância para a socialização do conhecimento.

Diante da dinâmica complexa do conhecimento implícito, as estruturas estáticas de KOS da organização do conhecimento podem ficar descontextualizadas por ser fruto de uma concepção baseada num contexto específico, numa certa configuração dimensional de tempo e espaço, a partir da perspectiva subjetiva de quem as criou e que, por isso, podem não mais serem adequadas para outro contexto. Segundo Souza:

Na tentativa de recuperar a perda de contextualização, diferentes mecanismos de reagrupamento foram desenvolvidos usando artifícios de diferentes naturezas para garantir maior significação na representação dos assuntos. (SOUZA, 2007, 111-112)

É importante considerar o grau de representatividade que os KOS desempenham em relação ao objeto referente. Para certas situações, alguns tipos de KOS podem ser mais adequados do que outros para melhor representar um objeto ou fenômeno.

Quanto maior a expressividade semântica e flexibilidade estrutural do KOS, mais fácil é a sua capacidade de representar melhor o conhecimento. As estruturas que melhor representam uma ampla gama de objetos da realidade são as ontologias

– um conjunto comum de termos e relações entre termos com elevada expressividade semântica, utilizadas para descrever e representar um domínio, uma área específica do conhecimento (YU, 2011, p. 137).

Com os KOS a organização do conhecimento procura reduzir a complexidade do conhecimento explícito por meio de sua redução a componentes mais elementares e distintos, os dados, e também em metadados (dados sobre dados), afim de obter maior controle de manipulação, acessibilidade, e outras capacidades que possibilitem o acesso e a utilização pelos usuários.

Por meio dos KOS, a organização do conhecimento faz a representação mediadora dos fenômenos a partir de dados e metadados. Ou seja, considera os dados como unidade de conhecimento e representação. Por outro lado, existe a abordagem que considera um “conceito” como a menor unidade de conhecimento e representação.

Conceitos são abstrações mais ricas em termos semânticos do que os dados e permitem uma representatividade mais próxima da realidade. Porém, exigem uma definição formal mais elaborada sendo, por sua constituição mais complicada, uma abordagem de representação mais apropriada para processamento por computadores.

## **2.7 Representação do conhecimento**

A representação do conhecimento (RC) é uma disciplina que se preocupa com a melhor representatividade dos fenômenos considerando a natureza dinâmica do conhecimento e a complexidade dos processos conversão entre conhecimento implícito e conhecimento explícito (socialização, externalização, internalização e combinação).

A representação do conhecimento é um processo que visa reproduzir as características de um dado fenômeno, procurando compreender, apreender, e representar o que são as coisas, como são formadas, como existem, como podem ser referenciadas e como se relacionam com outras coisas.

É importante ressaltar que é necessário definir antes o que se quer

representar.

Na filosofia clássica, a busca pela definição dos fenômenos (aquilo que se manifesta) chama-se “ontologia” - uma palavra derivada de outras duas, de origem grega: “*ontos*” e “*logos*”, que juntas significam “conhecimento do ser”, ou seja, o estudo das características e propriedades do ser como ele é, como se apresenta, como se manifesta. (CUNHA; CAVALCANTI, 2008, p. 268)

Definir o “ser” não é uma atividade trivial, haja vista a dificuldade de representar os aspectos dinâmicos e complexos que podem ser encontrados nos fenômenos. Gómez-Pérez et al. (2004, p. 2) levantam a questão da representação do conceito de “semente” e “árvore” em um banco de dados e questiona: uma semente é uma árvore? A resposta depende de como é modelada esta representação de maneira que possa expressar diferentes pontos de vista, considerando a dinâmica do contexto, da interpretação e da dimensão temporal das demandas.

Um outro exemplo: conforme mostra a figura 5, num dado momento “M1” existe uma propriedade rural de nome “A”. Se em outro momento “M2” esta propriedade rural mudar seu nome para “B” e não houver uma representação adequada que faça referência ao antigo nome “A”, o sistema de informação envolvido pode considerar, numa busca de dados, a não-existência de “A”, perdendo conexão com passado, e considerar apenas a existência de “B” como sendo “outra” propriedade rural, apesar das características físicas da área e da localização serem idênticas em “A” e “B”. E se, em momentos “M3” e “M4” a propriedade “B” for incorporada juntamente com uma propriedade “C”, e se a representação basear na simples descrição nominal de “B” e “C”, perde-se a identidade de “B” em “M4” em virtude da incorporação da área de “B” por “C”. Ou seja, “B” deixa de existir para se tornar “C”. Com representações baseadas em descrições nominais ou sem referências a informações históricas, não há como consolidar uma integração de dados que envolvam “A” nos momentos de “M1” a “M4”, pois a identidade e a definição do que seja “A” se perde no processo. Por outro lado, se a representação das propriedades rurais levarem em conta dados minuciosos sobre as posições georreferenciadas de “A”, “B” e “C” (além dos dados descritivos), é possível recuperar, sobre a dimensão geográfica, a definição de “A” de “M1” a “M4”,

independente do fato de haver fusão ou de desmembramento das propriedades rurais.

A figura 5 mostra as transformações da área de “A” para área “C”, denotando, em “M4”, que a definição geográfica de “A” se mantém por meio de informações georreferenciadas, mesmo se “A” perder sua referência nominal ou outras que poderiam identificá-la ou distingui-la dentro de “C”.

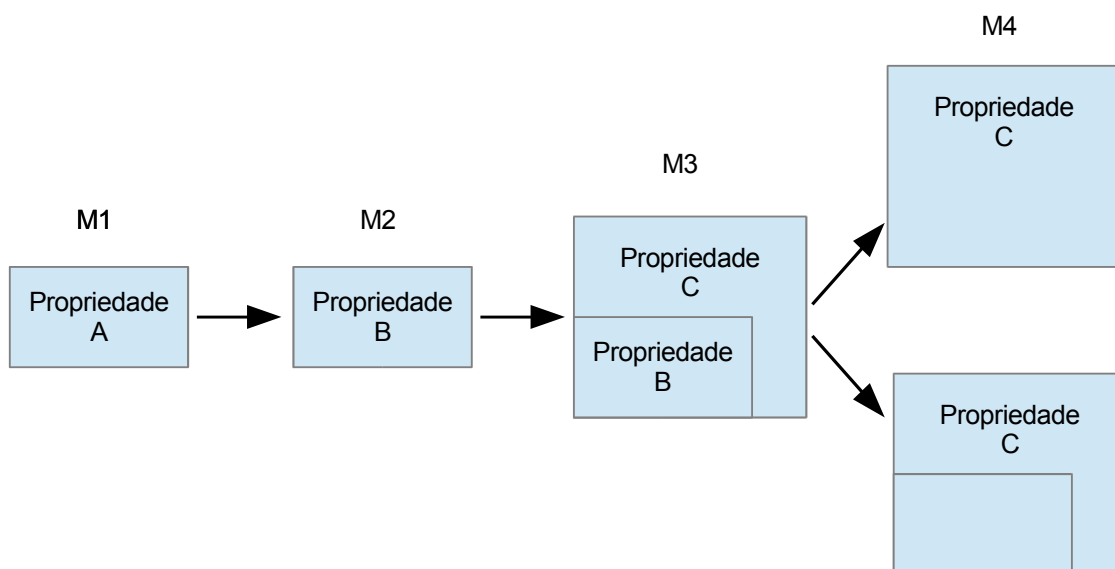


Figura 5 – Evolução de propriedades rurais.

Todavia, a conceitualização se destaca como processo de definição dos fenômenos. O conhecimento pode ser organizado naturalmente em conceitos de forma tanto implícita quanto explícita.

A conceitualização é um processo natural do ser humano em formular, por meio da linguagem, seus pensamentos em relação as coisas do mundo exterior. Os conceitos são abstrações caracterizam e definem os objetos que podem ser distinguidos. (DAHLBERG, 1978)

Genesereth e Nilsson (1987, apud Gruber, 1993, p. 1) afirmam que a representação do conhecimento se baseia em conceituações: objetos, conceitos, outras entidades e suas relações que denotam alguma área de interesse e que compreendem uma visão abstrata e simplificada do mundo que se deseja representar com algum propósito.

Conceitos são as unidades do conhecimento que se pretende expressar: “é a síntese de características necessárias que podem ser estabelecidas sobre qualquer referencial, representado por significações (termos, nomes ou códigos)”. (DAHLBERG, 2006)

De acordo com Guarino et al. (2009, p. 15), o processo de interpretação das coisas por meio de conceitos pode ser compreendido pela relação que produz significado na mente envolvendo signo, conceito e coisa. Esta relação pode ser representada por um esquema gráfico chamado Triângulo Semântico (figura 6). Neste esquema, a mente percebe um fenômeno externo (coisa) por meio de conceitos. Um conceito é formado dentro de um contexto restrito. Um termo ou palavra (signo) invoca o conceito e, desta forma, faz referência indireta ao fenômeno.

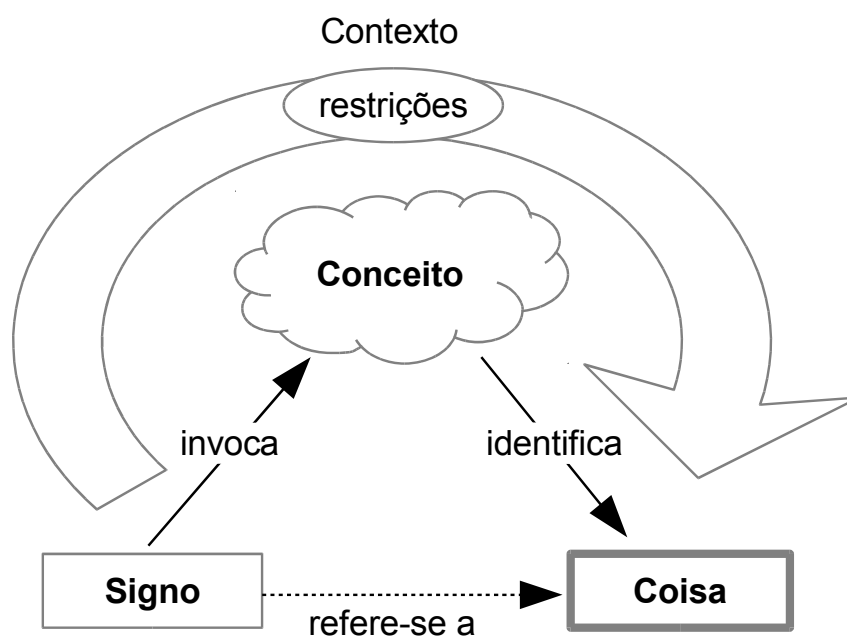


Figura 6 – Triângulo Semântico. (GUARINO et al., 2009, p. 15)

O processo de conceitualização envolve a análise dos conceitos e a síntese de suas características (DAHLBERG, 1978). Pode ser também compreendido, segundo Guarino et al. (2009, p. 8-10), como um processo que se inicia com a percepção dos fenômenos por meio da identificação de padrões, das características

invariantes (essenciais) e relevantes. Além disso, a conceitualização envolve outros dois aspectos:

- a) é um processo de abstração, pois as características muito detalhadas ou complicadas, assim como características irrelevantes, não são aproveitadas;
- b) é intermediado pela linguagem, onde a especificação dos conceitos pode ser feita de maneira intensional e extensional.

Para Dahlberg (1978), a análise dos conceitos parte do estudo de enunciados que fazem referência a uma ideia comum, independente de haver ou não um nome que a identifique. A partir da análise dos sujeitos e predicados dos enunciados, é possível distinguir os primeiros conceitos, suas características. A síntese destas características consolida o conceito em torno de uma definição mais formal, ajuda na formação de um nome para os conceitos que não possuem esta referência de identidade, e ajuda uma classificação dos conceitos.

Neste sentido, Guarino et al. (2009, p. 8) explica que o uso intensional da linguagem na formulação de conceitos visa restringir as interpretações e as ambiguidades produzidas pela linguagem por meio de axiomas adequados.

Partindo da análise intensional pode-se desenvolver a análise extensional por meio do processo de classificação, onde os conceitos podem ser agrupados com base nas propriedades comuns ou estratificá-los com base nos diferentes níveis de abstração, com conceitos mais gerais no maior nível de abstração e abrangência de definição, e conceitos individuais, mais específicos, em nível de maior detalhamento. Posteriormente, análise extensional prossegue com a avaliação das possíveis relações entre conceitos, bem como das possíveis definições de um mesmo conceito em no contexto de várias dimensões (como tempo e espaço).

Todo este processo pode ser revisto e refinado em várias iterações até que os conceitos tenham suas definições, características e relações consolidadas.

A conceitualização tem diversas abordagens de implementação. A mente humana faz conceitualização de uma forma tácita, sem a necessidade de alguma heurística bem definida. Os trabalhos de Dahlberg (1978, 2006) destinam-se



predominantemente à organização do conhecimento, particularmente os estudos sobre classificação. Os estudos de Gruber (1995), Sowa (2000b) e Guarino (2009) focam a conceitualização mais formal, voltada para a representação do conhecimento com bases em pressupostos lógicos e matemáticos.

Historicamente a representação do conhecimento surgiu das reflexões sobre a natureza das coisas, dos seres. Dentre esses estudos, destacam-se os de Aristóteles, sobre a organização do conhecimento baseada numa classificação geral das coisas existentes, as categorias. Com o desenvolvimento dos estudos sobre o raciocínio formal a partir da matemática e da lógica, o termo “ontologia” passou a definir as coisas na perspectiva moderna de conceitualização formal.

Como mecanismo de representar formalmente muitos tipos de conhecimento, as modernas ontologias estão presentes em diversas áreas de investigação (GUARINO, 1998), como por exemplo:

- a) engenharia de conhecimento;
- b) representação do conhecimento;
- c) modelagem qualitativa;
- d) engenharia de linguagem;
- e) projeto de banco de dados;
- f) modelagem da informação;
- g) integração de informações;
- h) análise orientada a objetos
- i) recuperação e extração de informações,;
- j) gestão do conhecimento;
- k) organização do conhecimento;
- l) projeto de sistemas baseados em agentes;
- m) integração da informação na empresa (*integration enterprise*);
- n) tradução de linguagem natural;
- o) medicina;
- p) engenharia mecânica;
- q) padronização do produto do conhecimento;
- r) comércio eletrônico (transações, etc);

s) sistemas de informações geográficas, legais, e biológicas.

Gruber (1993; 1995) define o sentido moderno de ontologia ao conceituá-la como uma especificação explícita de uma conceitualização.

Para Benjamins et al. cada elemento do conceito de Gruber significa:

"Conceitualização" refere-se a um modelo abstrato de algum fenômeno no mundo por ter identificado os conceitos relevantes do fenômeno.

"Explícito" significa que o tipo de conceitos usados e as restrições sobre seu uso são explicitamente definidas.

"Formal" refere-se ao fato de que a ontologia deve ser legível por máquina.

"Compartilhada" reflete a noção de que uma ontologia captura conhecimento consensual, isto é, ele não é privado de um indivíduo, mas aceito por um grupo. (BENJAMINS et al., 1998, p. 5-2)

Guarino (1997, p. 143) ressalta que ontologia é uma representação que restringe uma conceitualização particular, sem ser uma especificação direta, mas uma "caracterização axiomática do significado de um vocabulário lógico." E, complementa: "em muitos casos, os axiomas de uma ontologia apenas expressam subsunções (relações do tipo "é-um") e relações entre predicados unários. Porém, é necessária uma axiomatização mais detalhada para se excluir interpretações indesejadas."

Sowa (2000a, xi) considera a representação do conhecimento baseada em três componentes:

- a) a lógica, que provê a estrutura formal e regras de inferência como modelo de raciocínio;
- b) a ontologia, que define os tipos de coisas que existem num domínio de aplicação; e
- c) a computação, que atua como suporte que possibilita a implementação do modelo de inteligência artificial.

A representação do conhecimento dos filósofos da Antiguidade se baseava em duas áreas: lógica e ontologia. Na representação do conhecimento moderna esses dois conceitos se juntam com a computação, sinalizando a possibilidade de operação dos modelos de representação e inferência fora da mente humana, permitindo a reprodução, o compartilhamento, e processamento do conhecimento

explícito. Porém, a moderna RC não é capaz de reproduzir a geração de conhecimento e inteligência, qualidades típicas dos seres humanos.

A disciplina de representação do conhecimento também procura reproduzir, em sistemas computacionais, tarefas que normalmente são realizados pela inteligência humana (SOWA, 2000a, p. xi; VICKERY; VICKERY, 2004, p. 158).

Neste sentido Kashyap et al. relatam os avanços da ciência da computação:

Pesquisadores da área de representação do conhecimento desenvolveram esquemas de conhecimento expressivas de representação e teorias para capturar semântica da informação e processos de forma declarativa. Esta comunidade tem-se concentrado na implementação de mecanismos de raciocínio e de inferência para verificar a validade e satisfação das especificações do conhecimento. (KASHYAP et al., 2008, p. 3)

Nas aplicações tradicionais de inteligência artificial, nos programas de computador, nos esquemas de banco de dados, nas estruturas de arquivos, o conhecimento representado está inserido ou misturado em seu código interno. Porém, as ontologias são conhecimento representado de forma independente de aplicação específica. Neste sentido, Pospescu e Xu (2009, p. 3) afirmam que “o conhecimento de um domínio pode ser claramente explicado e compartilhado entre diferentes aplicações computacionais.”

Ontologia é um recurso que facilita o compartilhamento e reuso de conhecimento, uma forma de caracterizar o significado (semântica) das informações, sendo qualificada como uma especificação formal de representação do conhecimento, ou seja, que possui um grau de rigor lógico, sistemático, podendo ser codificada para uma máquina computacional. (GUARINO, 1997, p. 140; FENSEL, 2004, p. 3-10)

Segundo Lassila e McGuinness (2001), as ontologias podem variar em um “espectro ontológico” de menor para maior grau de representação e significação em relação ao objeto representado (expressividade semântica). Gomez-Perez et al. (2004, p. 28) destacam as principais categorias:

- a) vocabulários controlados: lista finita de termos referentes a coisas de um dado domínio de conhecimento;
- b) glossários: lista de termos com descrições (significados);
- c) taxonomias: lista de termos cujos termos estejam dispostos numa relação

- hierárquica, onde os termos mais específicos (hipônimos) possuem uma relação de dependência com um termo com significado mais genérico (hiperônimo)<sup>6</sup>;
- d) hierarquia informal: lista com termos dispostos numa relação hierárquica que não seguem necessariamente uma relação de significância mais geral para mais específica;
- e) tesouros: lista com termos e suas descrições ligadas a relações com outros termos. Essas relações podem ser do tipo hierárquico, associativo ou de equivalência (sinônimos);
- f) *frames*: modelos que incluem classes (*frames*) e propriedades, também chamadas de atributos (*slots*);
- g) ontologias de restrição de valor; e
- h) ontologias que exprimem restrições lógicas mais elaboradas (disjunção, inversão, parte-de, etc).

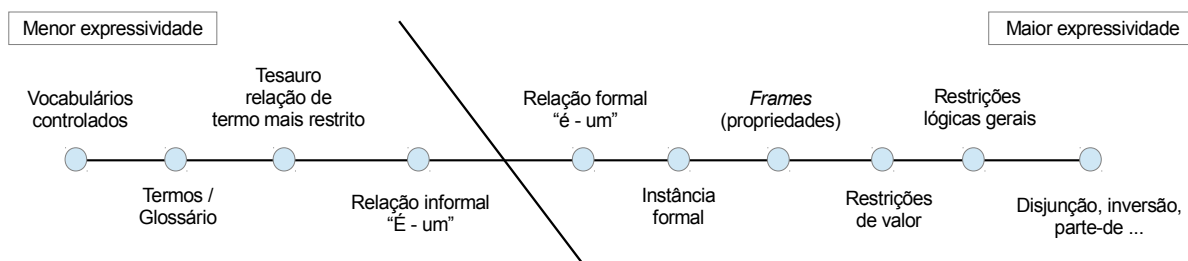


Figura 7 - Espectro ontológico (categorização) de Lassila e McGuinness descrito por Gómez-Pérez et al. (2004, p. 28)

Lassila e McGuinness consideram os esquemas de KOS e as ontologias como elementos de um único espectro ontológico (Figura 7). Nesse espectro, mais à esquerda situam-se as ontologias de “peso-leve”, ou seja, os esquemas de KOS (listas, taxonomias, tesouros, etc.) como tipos de menor expressividade semântica, com relações entre conceitos mais simples e, pela direita, as “ontologias de maior

6 Hiperonímia e hiponímia são tipos de relações semânticas da Semântica Lexical, um campo de estudo sobre como as palavras de uma língua denotam significados. O Apêndice B mostra um mapa conceitual com os tipos de relações semânticas.

peso semântico, como os esquemas com restrições lógicas gerais e relações mais elaboradas, de maior expressividade.

Os esquemas de KOS e ontologias são estruturas de representação do conhecimento formadas por conceitos e relações entre conceitos. Estas relações também são chamadas de relações semânticas (Apêndice B), pois quando dois ou mais conceitos se conectam é possível expressar um significado, uma semântica. Hjørland (2007b) elenca alguns tipos importantes de relações semânticas entre dois conceitos, dentre os quais destacam-se:

- a) relação ativa: quando expressa uma operação ou processo de um conceito afetar o outro;
- b) relação associativa: uma associação simples, sem especificidade;
- c) relação causal: quando um conceito A causa algo sobre um conceito B;
- d) homônimo: dois conceitos A e B são expressos pelo mesmo símbolo;
- e) hipônimo / hiperônimo: uma relação hierárquica onde um conceito (hipônimo) está subordinado a outro conceito (hiperônimo). Essa relação hierárquica ainda pode ser dividida em:
  - relação “é-um” ou “tipo-de”: esta relação denota que um conceito é membro do grupo de outro conceito. Quando todas as relações de um sistema semântico de conceitos é do tipo “é-um”, o sistema é uma taxonomia;
  - relação “específico-de”: permite indicar o conceito mais geral (hiperônimo) de todos os conceitos particulares;
- f) “instância-de”: denota a relação entre um conceito de abstração superior com outro conceito de abstração inferior (Cidade-Rio de Janeiro);
- g) relação locativa: indica que a localização de um conceito é designada por outro conceito;
- h) merônimo / holônimo : relação partitiva, do tipo “parte-de”, indica uma composição entre conceitos, onde um conceito relaciona-se como parte de outro conceito;
- i) relação passiva: ocorre quando um conceito é afetado ou é submetido à ação de outro conceito;

- j) relação paradigmática. De acordo com Wellisch (2000, p. 50 apud HJORLAND, 2007b), esta relação representa uma conexão fixa por natureza, auto-evidente, ou estabelecida por convenção (mãe-filho, gordura-obesidade);
- k) polissemia: ocorre quando uma palavra que tem vários sub-sentidos que estão relacionados um com o outro;
- l) antônimo: o significado do conceito A é oposto do significado do conceito B;
- m) sinônimo: conceito A tem o mesmo significado que o conceito B, pode ser expressa por “equivalente-de”.

Alguns autores classificam as ontologias pelo seu grau de expressividade e nível de complexidade: os esquemas de KOS, com relações semânticas e propriedades que descrevem os conceitos, caracterizam as chamadas ontologias de “peso-leve”; por outro lado, as ontologias “peso-pesado” adicionam maior expressividade semântica por meio de axiomas e restrições aos componentes de ontologia de “peso-leve”. (GÓMEZ-PÉREZ et al., 2004, p. 8)

Dadas essas características e capacidades de representação do conhecimento, no âmbito de uma organização, uma base de conhecimento central, estruturada e integrada por meio ontologias, pode se tornar um componente fundamental para captura do conhecimento explícito corporativo, reuso dos conhecimentos acumulados, e descoberta de novos conhecimentos.

## **2.8 Sistema de informação baseado em ontologias**

De acordo com Sowa (2000a, p. 417-419), os conhecimentos incorporados nos sistemas de informação estão codificados em cada programa, em cada repositório de dados, arquivos ou bancos de dados, muitas vezes em linguagens diferentes e diferentes formatos, desenvolvidas por pessoas diferentes, operando de acordo com diferentes padrões, contribuindo para a formação de um contexto de redundância e incompatibilidade da informação.

Nestas condições, o fluxo da informação, bem com a geração de

conhecimento, podem ser prejudicados em função desta heterogeneidade arquitetural.

Sowa (2000a, p. 418) apresenta um modelo de sistemas de informação com base em três componentes: banco de dados; aplicativos; e interface com o usuário. Cada um desses possui o conhecimento distribuído em seus respectivos esquemas conceituais, atuando de forma independente e semelhante às ontologias: os esquemas de bancos de dados definindo sua estrutura interna de armazenamento e acesso aos dados<sup>7</sup>; os aplicativos com suas lógicas internas codificadas; e a interface do usuário com o seu vocabulário, normas e protocolos específicos.

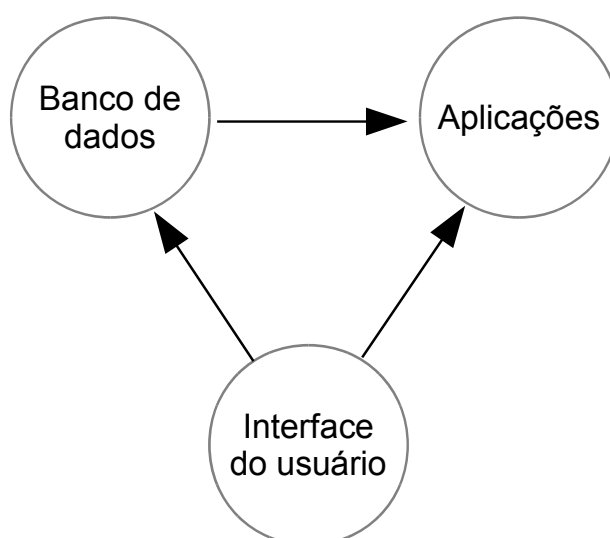


Figura 8 - Interfaces nos sistemas atuais (SOWA, 2000a, p. 418).

Por outro lado, um esquema conceitual central para todo sistema de informação pode aumentar a qualidade da geração, uso e compartilhamento de conhecimento (figura 8), integrando de forma homogênea as informações e operando a partir de um único padrão de linguagem e vocabulário.

---

7 "de fato, uma ontologia pode ser comparada com um esquema de banco de dados" (GAURINO, 1998, p. 10)

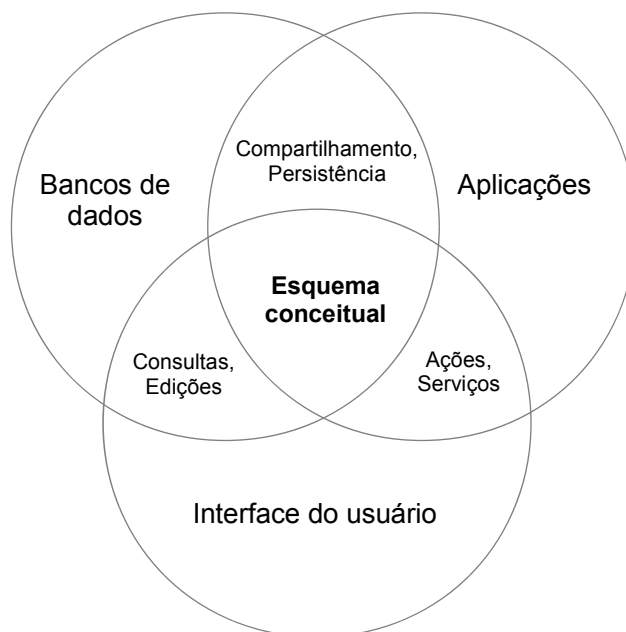


Figura 9 - Esquema conceitual como o coração de um sistema integrado (SOWA, 2000a, p. 418)

De acordo com Yu (2011, p. 1), integração de dados é o processo de combinação e agregação de recursos de informação para serem utilizados de forma coletiva.

No âmbito institucional, se os dados não estiverem integrados, eles não podem ser utilizados numa visão corporativa. É difícil de lidar com um ambiente corporativo onde os dados não estão integrados (INMON, 2002, p. 19).

É preciso integrar informações e estruturas, que estão normalmente fragmentadas em definições de metadados separados, em aplicações especialistas, específicas para um determinado domínio de aplicação, com códigos escritos em linguagem específica, mantidos por técnicos especializados.

As informações em diferentes fontes podem ser integradas e combinadas de forma a prover ao usuário uma visão unificada desses dados (LENZERINI, 2002, p. 233).

A integração das fontes pode ser realizada por um mecanismo lógico que provê uma padronização, estruturação e interface uniformes da informação acessada (LEVY, 2000, p. 3-4; 6). Além disso, a semântica dos dados pode ser



capturada por algum tipo de mapeamento semântico, que pode ser uma simples associação de termos para termos ou, em outros casos, uma fórmula complexa, muitas vezes expressa na lógica, linguagem de consulta, ou ontologia. (AN et al., 2006, p. 1)

Porém, a integração não é uma mera agregação, ligação, ou associação de recursos informacionais. É preciso que seja estabelecida uma arquitetura de integração que promova a emergência de novos significados, novos conhecimentos. Neste sentido, Sowa argumenta que o alto grau de estruturação de uma organização (arquitetura) é a parte fundamental dos sistemas de informação, independente dos computadores:

O principal requisito para um sistema de banco de dados não é um computador, mas uma organização altamente estruturada, necessária para armazenar e processar grandes volumes de dados. As organizações empresariais são baseadas em princípios de manutenção de registros e de contabilidade, que foram desenvolvidas e aperfeiçoadas desde os tempos dos antigos sumérios. Quando os computadores se tornaram disponíveis, as empresas tinham estruturas de registro bem definidas que podem de imediato ser mapeados para uma forma computável.

[...] sistemas simples em pequenos computadores [...] sistemas maiores [...] O aspecto comum que unifica todos os grupos é um conhecimento do significado dos dados e as restrições necessárias para manter o modelo fiel com o mundo real. (SOWA, 1984, p. 303-304)

O importante é saber o significado de um pedaço de informação, ou quais entidades fazem referência a ela. Para Levy, o processo de integração já incorpora um certo grau de significância:

um sistema de integração de dados requer um mecanismo flexível para descrever o conteúdo de fontes que podem ter conteúdos sobrepostos, cujos conteúdos são descritos por restrições complexas, e de fontes que podem ser incompletas ou apenas parcialmente completas. (LEVY, 2000, p. 2).

Vale ressaltar que a riqueza de um repositório central de dados reside não apenas na integração dos dados e sua acessibilidade, mas também na capacidade de processar e avaliar os significados dos dados acessados. Mesmo integrados, em muitas situações de recuperação e uso de informações, os repositórios têm um papel passivo como meros depósitos de informação, onde as inferências ainda são processos intrínsecos da mente do usuário.

Por isso, a arquitetura de integração de dados ideal é aquela capaz de

colaborar efetivamente para geração novos conhecimentos, não como um banco de dados, ou depósito de dados, mas como uma base que produz conhecimento.

Uma base de conhecimento (KB, *knowledge-base*) é um tipo de banco de dados destinado a armazenar não apenas registros de dados, mas também integrar todo o tipo de informação considerado como conhecimento explícito no âmbito de uma organização social afim de produzir novos conhecimentos e promover inovação.

Uma base de conhecimento é mais do que um armazém que converge, registra, integra e mostra dados. Trata-se de um repositório capaz de “descrever fatos e afirmações relacionadas a um determinado estado de coisas ou um estado epistêmico particular.” (GUARINO, 1997, p. 144)

No modelo de esquema conceitual integrado, Sowa (2000a, p. 495) considera base de conhecimento como “uma coleção de informação que inclui uma ontologia como componente”.

Numa arquitetura de base de conhecimento, as ontologias podem representar melhor o conhecimento do que os bancos de dados relacionais. Uschold (2006) destaca alguns aspectos comparativos a favor das ontologias:

Quadro 1 – Comparação entre banco de dados e ontologias

<b>Esquema de banco de dados</b>	<b>Ontologia</b>
Foco nos dados.	Foco no significado.
Restrições de banco de dados: - para garantir a integridade; - pode sugerir o que significa.	Axiomas ontológicos: - para especificar significado; - pode ser usado para integridade.
Não possui estrutura hierárquica	Hierarquia é sua estrutura principal
Utiliza SQL para: - visões e consultas; - integridade dos dados.	Utiliza mecanismos de raciocínio: - para inferir novas informações; - garantir a coerência.
Centrado nas instâncias.	Instâncias são opcionais.
Dicionário de dados é um artefato separado.	Comentários podem ser parte da teoria.

Fonte: Uschold (2006).

Segundo Guarino (1997, p. 145; 1998, p. 7), de acordo com o nível de dependência, as ontologias pode ser classificadas em quatro tipos (figura 10):

ontologias de alto nível; ontologia de domínio; ontologia de tarefa (ou de atividades); ontologia de aplicação.

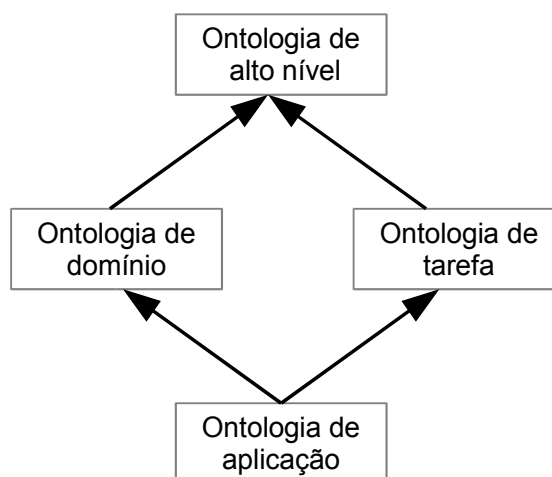


Figura 10 - Tipos de ontologias de acordo com o nível de dependência (GUARINO, 1997, p. 145)

Uma ontologia de alto nível é responsável pela integração de elementos de um sistema de informação, em um nível de abstração superior, sem. Ela descreve conceitos mais gerais como tempo, espaço, matéria, objeto.

Ontologias de domínio são extensas quanto à quantidade de definições de conceitos, descrevendo todos os conceitos e relações envolvidos no escopo da aplicação (seu domínio). Em outras palavras, as ontologias de domínio constituem os vocabulários definidos para a aplicação.

As ontologias de tarefas expressam os processos e suas atividades, como os passos lógicos devem ser executados.

Ontologia de aplicação são combinações de aspectos das ontologias de domínio com aspectos das ontologias de tarefas, podendo representar as atividades de alguma entidade do domínio.

Para BLOEHDORN et al. (2009, p. 4) ontologias podem auxiliar na integração de fontes de conhecimento heterogêneas, que podem ser estruturadas, semiestruturadas ou não-estruturadas. Além disso, o conteúdo podem abrigar conceitos semanticamente iguais, porém com representação distinta.

Neste sentido, em um sistema de informação com base de conhecimento de ontologias pode haver esquemas de integração baseados em ontologias de alto nível conectados à diversas fontes de dados heterogêneas, além de programas de aplicação do sistema de informação ou módulo de interface com de parte de seu código estático conectado a uma ontologia de tarefa ou ontologia de aplicação<sup>8</sup>.

Com base nestas características, é possível deduzir um sistema de informação composto pela combinação do esquema conceitual do modelo de integração de três elementos de Sowa com os quatro tipos de ontologias de Guarino (figura 11).

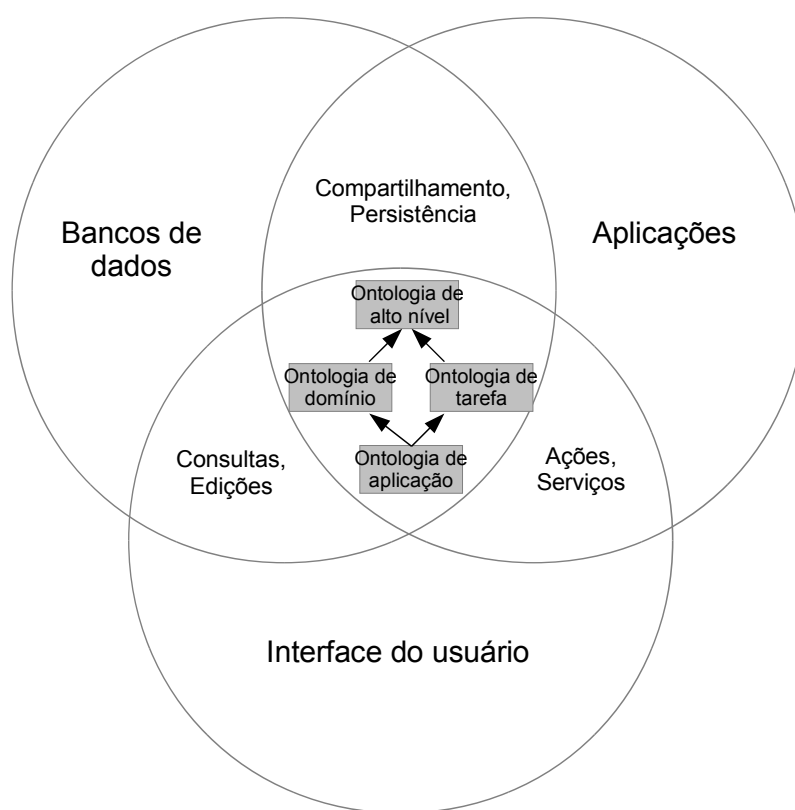


Figura 11 - Esquema conceitual de integração constituído por ontologias, adaptado de Guarino (1997) e Sowa (2000a)

<sup>8</sup> Uma visão de implementação deste modelo pode ser vista em Hebel et al. (2009, p. 174; 301-304; 468).

Um sistema de informação com base de conhecimento de ontologias pode ter grande impacto favorável para a gestão do conhecimento corporativo. Dieng-Kuntz (2005, p. 67-80) mostra que, apesar da ênfase das ciências da gestão nesta área, pesquisas da área de inteligência artificial e sistemas especialistas evoluíram para a idealização de sistemas baseados no conhecimento, caracterizados pelas bases com ontologias que, por suas características, constituem a memória da organização (em termos de conhecimento explícito), da qual podem emergir novos conhecimentos, apoiados pelos mecanismos de inferência sobre as ontologias.

Com relação às ontologias, estas são compostas por classes, propriedades componentes das classes, relações entre classes, especificações de funções e regras, e instâncias de classes. As classes representam uma abstração das instâncias, ou seja, são constituídas por elementos presentes em todas as instâncias que, por sua vez, são representações dos objetos que podem ser descritos e distinguidos de forma singular. Numa ontologia de nível mais alto, as classes representam conceitos mais gerais, de nível mais abstrato. As relações entre as classes podem ser expressas por hierarquias taxonômicas e meronômicas. Uma taxonomia é uma relação de generalização ou especialização de uma classe por outra classe. A subclasse herda as propriedades da classe mais geral, agregando novas propriedades. Uma meronímia é uma relação de composição onde uma classe é composta por uma ou mais classes. (GÓMEZ-PÉREZ et al., 2004, p. 6-8; BREITMAN, 2005, p. 30-42; 120; 126-137)

Uma base de conhecimento pode ser construída por meio de lógica descritiva (DL, *Description Logic*) - uma linguagem para descrição formal de ontologias que permite definir e descrever conceitos, relações semânticas, axiomas terminológicos ou assertivos (declarações e proposições ditas como certas e verdadeiras).

Segundo Gómez-Pérez et al. (2004, p. 17-20) e Baader et al. (2007), a DL permite a representação de ontologias com três tipos de componentes: conceitos (classes e objetos); papéis (relações binárias entre os conceitos); e indivíduos, que são as instâncias dos conceitos. A DL possui duas partes:

- a) parte intensional, terminológica, chamada de Tbox, que contém as definições dos conceitos e seus papéis, sob os quais podem operar um

- conjunto de construtos lógicos para a definição dos axiomas como conjunção, disjunção, negação, equivalência, restrição de valor, etc.; e
- b) parte extensional, assertiva, chamada de ABox, com as definições dos indivíduos, as instâncias dos conceitos, bem como os valores dos papéis (propriedades) dos conceitos.

Vale ressaltar que elementos desta linguagem de lógica descritiva possuem certa compatibilidade com elementos do processo de conceitualização proposto por Dahlberg (1978, 2006).

A composição de uma DL para Nebot e Berlanga (2012, p. 854) é um pouco diferente, mas relevante, em duas partes: Abox, que normalmente é muito dinâmico, constantemente atualizada; e Tbox, que dificilmente muda ao longo do tempo. Estas duas partes possuem os seguintes aspectos:

- a) parte terminológica (TBox), onde conceitos podem ser definidos em termos de outros conceitos ou de papéis que desempenha (propriedades e características) por meio de uma série de construtores (relações semânticas), a saber: união, intersecção e complemento, bem como enumerações, universalidade, cardinalidade, restrições sobre os papéis ou seu inverso. Definições conceituais são afirmadas como axiomas, que podem ser de dois tipos: subsunção<sup>9</sup> e equivalência. Ambas pode ser formuladas entre papéis de conceitos, e papéis podem ter restrições especiais (por exemplo, transitividade, simetria, funcionalidade, etc);
- b) parte assertiva (Abox), referente aos dados semânticos, que são expressos como afirmações individuais, que podem ser basicamente de dois tipos:
- um conceito de um indivíduo (uma instância);
  - uma relação entre indivíduos.

---

9 Uma definição extensional de subsunção caracteriza uma relação semântica do tipo “é-um”, que estabelece uma hierarquia entre conceitos – um superior (hiperônimo) e outro inferior (hipônimo), no sentido de que o conceito inferior herda as características do conceito superior e este, por sua vez, representa uma abstração maior. Ou seja, o conceito inferior é igual ao conceito superior, porém com características adicionais, com maior detalhamento (menor abstração). Por exemplo, na relação entre conceitos “carro” “é-um” “veículo”, os conceitos mediados pela relação “é-um” são hipônimo e hiperônimo, respectivamente.

Numa comparação, enquanto um banco de dados possui um esquema de definições e dados, em uma base de conhecimento com ontologia descrita em DL há um esquema de definições terminológicas (TBox), pouco mutável, às vezes considerado como a ontologia propriamente dita, e os dados semânticos (ABox), conforme o esquema da figura 12.

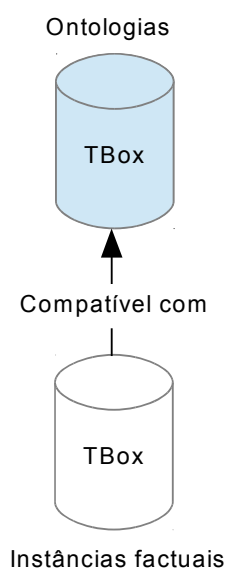


Figura 12 - Base de Conhecimento (LACY, 2005, p.26).

Uma base de conhecimento com ontologias e dados semânticos interligados pode ser mais apropriada para processamento de inferências, com buscas mais eficientes de informação do que a recuperação de documentos.

Por exemplo, ferramentas de busca na Internet conseguem recuperar como resultado (dentre outros) o termo "Titanic" a partir da entrada dos termos "Navio", "Iceberg" e "1912" porque, na grande base de conhecimento que é a Internet, estes quatro termos ("Navio", "Iceberg", "1912" e "Titanic" são termos que possuem alguma relação semântica já estabelecida pelos documentos onde estão contidos (arquivos de texto, páginas da Internet, etc). Ou seja, os documentos são a única garantia de relação semântica entre os termos.

A Internet é uma base de conhecimento de documentos, com semântica contida em seus conteúdos e, como já visto, agregar documentos não redundante na

agregação de informação para emergência de conhecimento. De acordo com Yu (2011, p. 333), páginas de conteúdo da web são destinadas para leitura pelos seres humanos. Seu conteúdo não é estruturado o suficiente para permitir o reuso automático do conhecimento (pelos computadores).

No processo tradicional de geração de conhecimento, com banco de dados sem ontologia, o usuário ou aplicação de busca automática deve intervir e elaborar heurísticas para lidar com documentos, não com elementos de informação. E para extrair conhecimento de banco de dados, ou qualquer outro conjunto de dados destituído de semântica própria, o usuário ou aplicação precisa estabelecer (ou codificar) as relações semânticas de acordo com suas necessidades e especificidades de contexto e domínio da aplicação.

Além da busca semântica, uma outra importante capacidade de uma base de conhecimento em DL é a inferência semântica<sup>10</sup>, que permite acessar a informação por meio indiretos, aproveitando as conexões dos elementos de informação.

O seguinte mapa conceitual demonstra um exemplo:

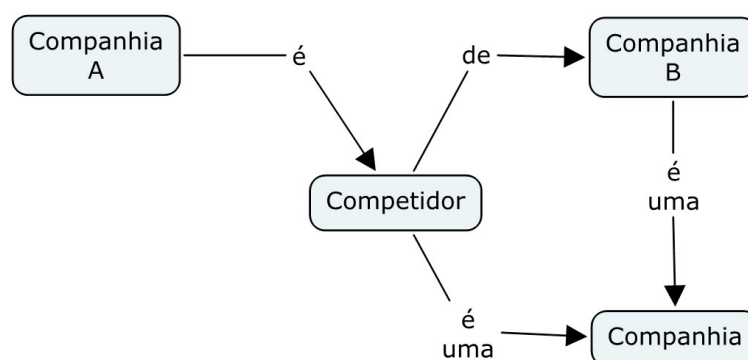


Figura 13 - Rede de conceitos para exemplo de inferência semântica, adaptado de Yu (2011, p. 356-357).

No caso da rede de conceitos da figura 13, uma requisição do tipo “retorne todas as companhias encontradas”, uma busca semântica retornaria “Companhia B”, pois esta instância (dado semântico, ou indivíduo em DL) está ligada por subsunção ao conceito “Companhia”. Por outro lado, numa inferência semântica, o resultado

<sup>10</sup> A implementação de um outro exemplo em computador utilizando recursos de Web Semântica pode ser vista em Hebel et al. (2009, p. 163-172).



seria “Companhia A e Companhia B”, pois de forma indireta, “Companhia A” também é uma companhia por intermédio do conceito “Competidor”.

No processo de reuso ou geração de novos conhecimentos a inferência semântica possui um papel importante. Porém, é preciso utilizar computadores nesta tal tarefa, que são capazes de armazenar grandes volumes de dados em pequenos espaços e tempos muito curtos.

Computadores podem separar e classificar dados porque possuem mecanismos básicos de ordenação que possibilitam determinar que o caractere “A” precede o caractere “B”. Mas esses mecanismos básicos não são suficientes para que a máquina compreenda o sentido da relação entre os conceitos “Inseto”, “Praga”, “Cigarrinha-verde” e “Feijão”. Segundo Nebot e Berlanga (2012, p. 853), é necessário adicionar entendimento às máquinas por meio de anotação semântica aos dados e adoção de ontologias, permitindo que computadores e pessoas possam recuperar e processar dados das mais variadas formas.

O carácter formal da ontologia é imprescindível para sua implementação em computador. De acordo com Sowa (2000a, p. 186), “para ser útil, um computador deve representar a informação sobre as coisas no mundo, mas a informação computadorizada passa por meio de muitos níveis de representação de representação de representação.”

Mas uma base de conhecimento em lógica descritiva não pode se inserida em computador, pois trata-se de uma linguagem que serve apenas para transformar o modelo conceitual (conceitualização) em especificação semi-computacional (GÓMEZ-PÉREZ et al., 2004, p. 110; BREITMAN, 2005, p. 74). Para tanto, é preciso implementar as ontologias em alguma linguagem inteligível por computadores. Enquanto o modelo do processo de conceitualização pode ser facilmente mapeável para o modelo formal de lógica descritiva (semi-computacional), esta linguagem última pode ser mapeável para o padrão tecnológico da Web Semântica (SW, *Semantic Web*).

## 2.9 Web semântica

A Web Semântica (WS) é uma tecnologia de infraestrutura destinada ao intercâmbio e integração de dados na Web, desenvolvida e mantida pelo Consórcio Internacional Gestor da Web (W3C, *World Wide Web Consortium*).

A proposta essencial para a Web Semântica é atribuir significado à estrutura de informação da Internet Web, pois, do ponto de vista das máquinas, tudo é visto como recursos (documentos) ligados a referências. Por outro lado, a Web Semântica agrega significado às estruturas e componentes da Web comum, indicando “o que” representa cada parte da estrutura (figura 14).

Ao falar sobre uma das linguagens da Web Semântica, Heath e Bizer (2011, p. 4), sintetizam as duas capacidades essenciais que transformam a Web em Web Semântica: as referências passam a ligar coisas (recursos) e não apenas documentos; e as ligações são tipadas, expressando o significado sobre as coisas que estão sendo ligadas.

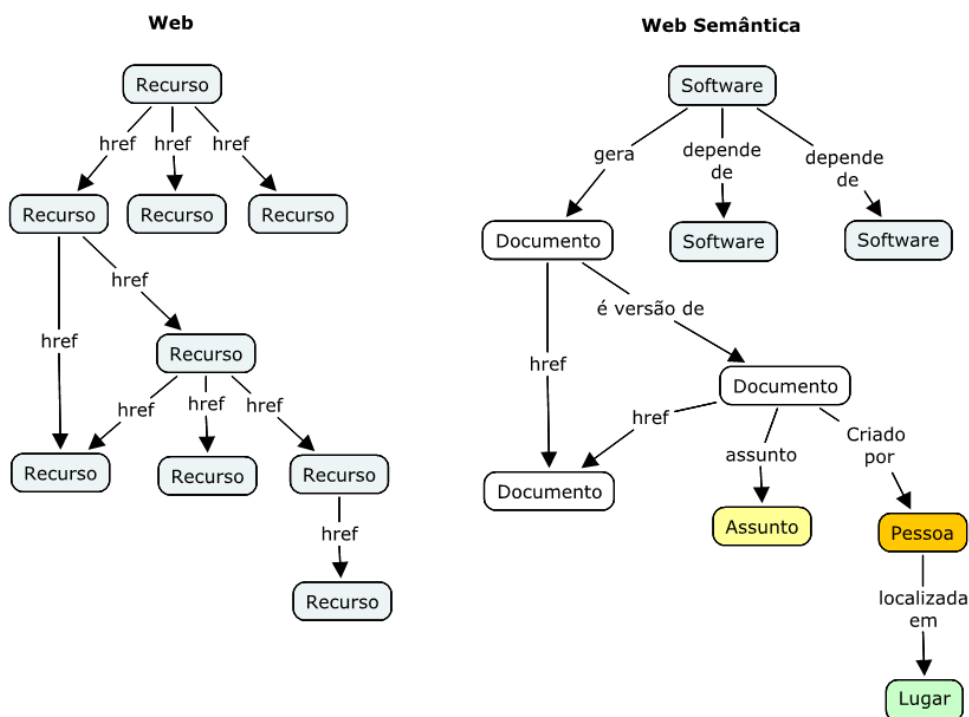


Figura 14 - Estrutura da Internet Web em comparação à estrutura da Web Semântica (W3C, 2001).

A WS atribui significado por meio de descrições nos dados e documentos, e representa o conhecimento por meio de ontologias. Sua arquitetura é disposta de forma escalável, desde as funções mais básicas e essenciais, até as operações mais elaboradas. Ou seja, a escalabilidade torna possível a expansão e o aumento das capacidades de uso da tecnologia sem a necessidade de alterar a constituição das estruturas e conteúdos de informação anteriores.

É constituída por uma coleção de padrões e tecnologias que possibilitam as máquinas entenderem o significado (semântica) da informação na Web, permitindo o compartilhamento e a reutilização de informações no âmbito de diferentes plataformas e aplicações de computadores. (YU, 2011, p. 15, 81)

A WS melhora a precisão e reduz a complexidade das pesquisas a dados e documentos, devido às ontologias poderem relacionar e associar informações por meio de processamento de inferências (BERNERS-LEE et al., 2001).

O conteúdo da WS é basicamente formado por dados, metadados e ontologias, localizados tanto em um ambiente de rede local quanto remotamente na Internet (W3C, 2005b). Os dados podem ser estruturados, semiestruturados ou não estruturados, além de estarem agregados a metadados, anotações e outras formas de descrição. Estas agregações podem estar presentes em documento ou até em dado atômico, a menor informação possível de ser recuperada. Por outro lado, as ontologias operam sobre os dados, possibilitando operações de raciocínio, inferência, integração, mediação e comunicação. (KASHYAP et al., 2008, p. 7-8; ANTONIOU et al., 2008, p. 8-12)

Um repositório de dados baseado em Web Semântica, pode ter uma camada integradora de dados, em linguagem padronizada, empregando ligações e hiperligações lógicas com os dados distribuídos. (HEATH; BIZER, 2011, p. 3)

### **2.9.1 A arquitetura da Web Semântica**

Um conjunto de recursos padronizados e escaláveis forma a arquitetura da WS, podendo ser representada como uma estrutura de camadas sobrepostas, uma

hierarquia de linguagens e especificações, denominada de “Pilha” da Web Semântica (em inglês, *Semantic Web Stack*). A figura 15 mostra a arquitetura original proposta pelo W3C, cujos componentes elencados de baixo para cima são os seguintes:

- a) URI/IRI - *Uniform Resource Identifier* (Identificador Uniforme de Recurso) e *Internationalized Resource Identifier* (Identificador internacional de recurso): faz um endereçamento (referência) indicando a localização dos recursos ou dados;
- b) XML - *eXtensible Markup Language* (Linguagem Extensível de Marcação): linguagem de base das outras linguagens de marcação da Web Semântica: RDF, OWL e RIF;
- c) RDF – *Resource Description Framework* (Conjunto de Classe de Descrição de Recurso): conjunto de especificações em XML que permite a ligação entre recursos ou dados;
- d) RDFS - *Resource Description Framework Schema* (Conjunto de Classe de Descrição de Recurso para Esquemas): conjunto de especificações ligadas que permite a construção de vocabulários;
- e) OWL - *Web Ontology Language* (Linguagem de Ontologias para a Web): esquema baseado em XML, RDF e RDFS que permite a construção de especificações semânticas de ligações entre recursos ou entre dados;
- d) SPARQL – *SPARQL Protocol and RDF Query Language* (SPARQL Linguagem de consulta e protocolo a dados em RDF): linguagem de consulta e protocolo de acesso a conjunto de dados semânticos (triplas);
- e) RIF – *Rule Interchange Format* (Formato para Intercâmbio de Regras): uma linguagem derivada da XML<sup>11</sup> para intercâmbio, entre aplicativos, de regras de negócios e diferentes estruturas de dados;
- f) Unifying Logic (Unificação Lógica): camada ainda em discussão que visa integrar logicamente as informações na Web;
- g) Crypto, Proof, Trust<sup>12</sup> (Criptografia, Prova, Confiança): são camadas em

---

11 Disponíveis em: [http://www.w3.org/2005/rules/wiki/RIF\\_FAQ#What\\_is\\_RIF.3F](http://www.w3.org/2005/rules/wiki/RIF_FAQ#What_is_RIF.3F) e <http://www.w3.org/TR/2013/NOTE-rif-primer-20130205/> (Acesso em: 6 fev. 2013)

12 Detalhes sobre confiança na Web Semântica estão disponíveis na página do W3C:

discussão que visam garantir segurança, veracidade e fidedignidade ao conteúdo publicado na Web. A ideia parte da premissa de que a informação deva ser verdadeira<sup>13</sup> e logicamente coerente.

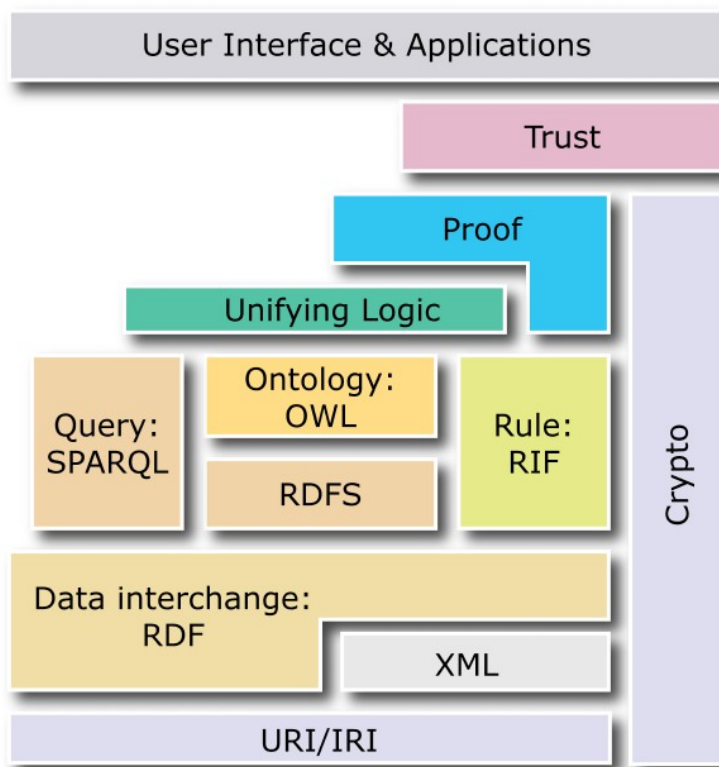


Figura 15 - Pilha da Web Semântica definida pelo W3C. (Fonte: W3C, 2007)

Do ponto de vista das aplicações, Lacy (2005, p. 44) propõe uma arquitetura mais simples, conforme a figura 16.

<http://www.w3.org/2000/10/swap/doc/Trust> . Acesso em: 6 fev. 2013.

13 Segundo Dretske (1999, p. 45), “a informação 'falsa' e meia-informação não são tipos de informações. Assim como patos de decoração e patos de borracha não são tipos de patos. [...] Informação é o que é capaz de produzir conhecimento, e uma vez que o conhecimento exige a verdade, a informação também a requer.”

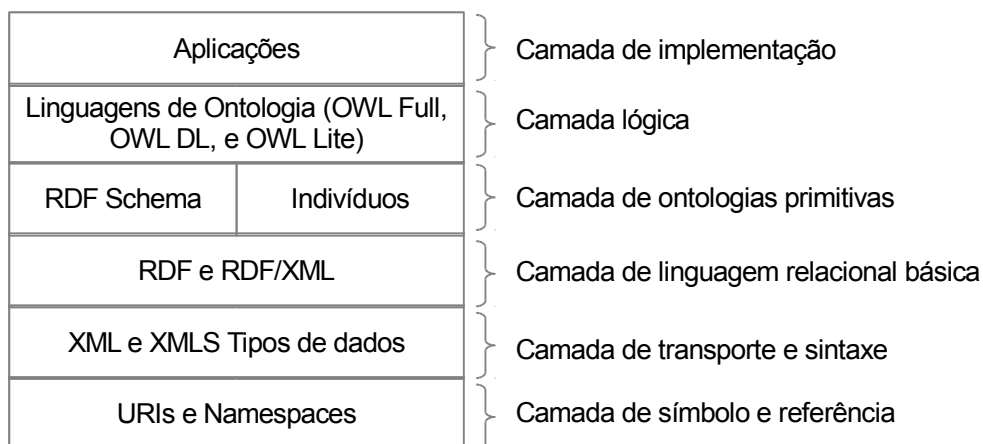


Figura 16 - Arquitetura de Camadas da Web Semântica (Lacy, 2005, p. 44).

De acordo com a arquitetura de Lacy (2005, p. 44), a pilha da Web Semântica possui na parte mais inferior os identificadores únicos de recursos - cadeias de caracteres que identificam nomes ou recursos na rede de computadores. O Identificador Uniforme de Recurso (URI) identifica o local e o nome do recurso na rede de computadores, e o Identificador de Recurso Internacional (IRI, uma variação do URI que aceita caracteres internacionais<sup>14</sup>.

Acima do URI/IRI encontram-se as linguagens derivadas da linguagem de marcação mais geral, denominada SGML (*Standard Generalized Markup Language*, Padrão Generalizado de Linguagem de Marcação). Estas linguagens são: XML, RDF, RDFS e OWL. Nesta sequência cada uma delas constituída por características derivadas da linguagem anterior, formando uma arquitetura escalável de linguagens.

Um exemplo de linguagem do tipo do padrão SGML amplamente utilizada é a Linguagem de Marcação em Hipertexto (HTML - *HyperText Markup Language*), que provê semântica de formatação padrão para documentos da Web, processados por programas de navegação na Internet. De acordo com Breitman (2005, p. 48), com a HTML “mudanças na apresentação da informação podem ser obtidas sem que seja necessário realizar mudanças no conteúdo dos documentos”. Porém, a semântica da HTML é específica para formatação de documentos, não sendo adequada para

<sup>14</sup> Sobre URIs e ligação de dados com Web Semântica ver o artigo do W3C: <http://www.w3.org/TR/cooluris/>. Acesso em: 11 fev. 2013.

representação semântica mais genérica.

Para maiores detalhes sobre a linguagem SGML e outras linguagens derivadas, consultar o mapa conceitual do Apêndice I.

Em um documento de linguagem de marcação um conjunto de elementos de informação em linguagem de marcação está contido numa estrutura hierárquica, onde cada elemento é delimitado por um par de marcadores (*Tags*) que lhe atribuem alguma semântica. Na primeira *Tag* de um par delimitador de marcação podem haver diversos atributos qualificadores que detalham mais a semântica do elemento marcado.

A linguagem HTML destina-se a atribuir semântica a elementos de apresentação de conteúdo de documentos para Web. Mas para estruturação e especificação de dados a linguagem mais adequada é a XML (*eXtensible Markup Language* - linguagem de marcação extensível).

Em XML é possível atribuir qualquer marcação a um dado conteúdo e, por isso, para garantir uma semântica mais formal e maior controle dos termos das marcações, é preciso estabelecer regras, restrições e pré-definição de *Tags*, por meio de especificações em arquivos de definição de tipo de documento (DTD, *Document Type Definition*).

Documentos de XML fazem referência a um DTD devem seguir as regras de marcação estabelecidas neste documento. E um documento XML pode fazer referência a diferentes documentos DTD herdando, assim, suas regras de especificação de *Tags*. Se em um ou mais DTD houver um termo de marcação semelhante, o mecanismo da XML denominado "*namespace*" resolverá as ambiguidades. No código XML o *namespace* está especificado por meio do atributo "*xmlns*" (*XML Name Space*), que fica situado geralmente no início do documento. Por outro lado, é importante ressaltar que, além das *Tags* herdadas e utilizadas a partir dos DTDs, o documento XML pode acrescentar suas próprias regras de especificação o permite uma escalabilidade entre os dados.

## 2.9.2 Resource Description Framework (RDF) e RDF Schema

Derivada da XML, a linguagem RDF (*Resource Description Framework*), é formada por um conjunto de definições de classes e propriedades em um vocabulário controlado bem simples, um conjunto de termos que é suficiente apenas para estabelecer uma ligação entre as informações. Por isso, foi desenvolvida a RDF Schema, ou RDFS, uma ampliação das capacidades do vocabulário da RDF<sup>15</sup>.

A RDFS é uma linguagem extensível de representação de conhecimento, útil para escrever vocabulários por meio de definição de classes e propriedades de um domínio específico. Uma classe é um recurso que compartilha características com outros recursos aos quais está associada. (YU, 2011, p. 111-114; W3C, 2004a; BREITMAN, 2005, p. 51; POWERS, 2003, p. 84)

Enquanto RDF é útil para estabelecer ligações, a RDFS permite criar vocabulários (esquemas), com definições de classes (conceitos), propriedades (características), bem como estabelecer relações hierárquicas como, por exemplo, de subsunção (“é-um”) por meio da propriedade “`rdfs:subClassOf`”<sup>16</sup>.

Em termos da ciência da complexidade, a linguagem RDF gera uma informação emergente, um significado, a partir da atribuição de sentido a uma ligação entre dois dados.

A partir de uma ligação, é possível estabelecer outras ligações, uma vez que os dados componentes podem se ligar a outros dados e estes a outros, e assim por diante, formando uma rede semântica, semelhante a conceitos interligados.

Segundo Yu (2011, p. 72), RDF representa e modela informação sob a forma de listas de declarações que podem ser processadas por computadores. Combinações de declarações RDF distribuídas podem ser agregadas para facilitar a descoberta de novas informações.

---

15 As definições do vocabulário RDF encontram-se no link: <http://www.w3.org/1999/02/22-rdf-syntax-ns>, e uma documentação encontra-se no link: <http://www.w3.org/TR/rdf-schema/>. Acesso em: 7 fev. 2013.

16 Esta propriedade de relação de subsunção está documentada no link: <http://semanticweb.org/wiki/Rdfs:subClassOf>. Acesso em: 7 fev. 2013.



O W3C (2004b), destaca na linguagem RDF possui as seguintes características:

- a) é uma projetada para representar informações que estejam isoladas, mas com capacidade de serem facilmente agregadas;
- b) permite estabelecer uma semântica formal sobre os dados e capacidade de inferência às aplicações;
- c) a estrutura básica de uma expressão em RDF é uma coleção de triplas, onde cada tripla é uma declaração de uma relação entre as coisas denotadas pelos nós que as vincula. Estes nós são três partes: um sujeito, um predicado e um objeto;
- d) o sentido da relação é sempre do sujeito para o objeto. O predicado sempre media o sujeito e o predicado;
- e) o sujeito pode ser uma referência de URI para um recurso; ou um nó em branco;
- f) nós em branco indicam um recurso anônimo, ou sua ausência. É um nó que estabelece uma ponte entre triplas sem a necessidade de ser definido;
- g) o predicado, também conhecido como "propriedade", somente pode ser uma referência à uma URI para um recurso;
- h) um objeto pode ser uma URI, um literal, ou um nó branco. Com exceção de ser um literal, um objeto pode ser sujeito de outra tripla e assim por diante (figuras 17 e 18).

Alguns exemplos de códigos escritos no formato RDF/XML estão descritos no Apêndice K, cuja validação pode ser feita por meio de um programa do W3C disponível na Internet, conforme descrito no Anexo F.

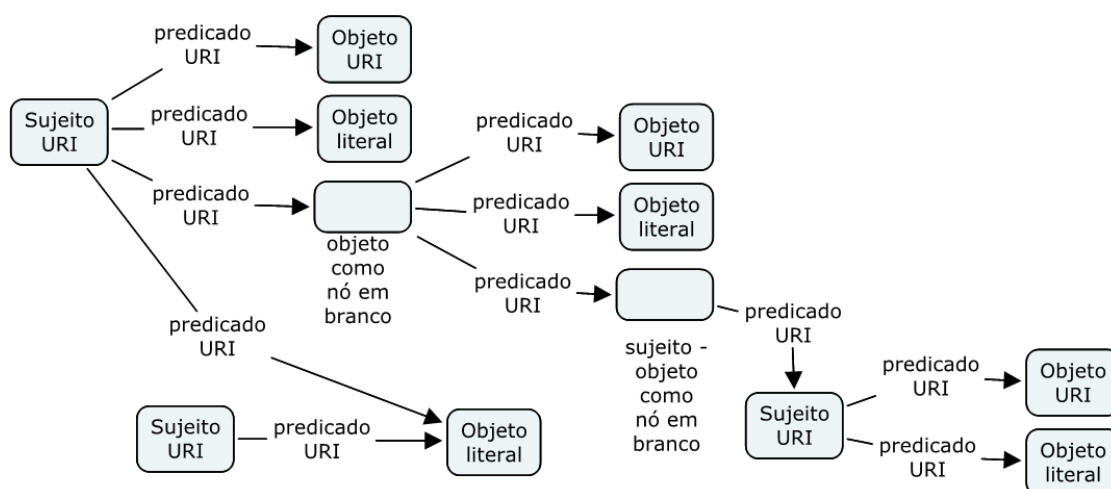


Figura 17 - Esquema conceitual de triplas RDF.

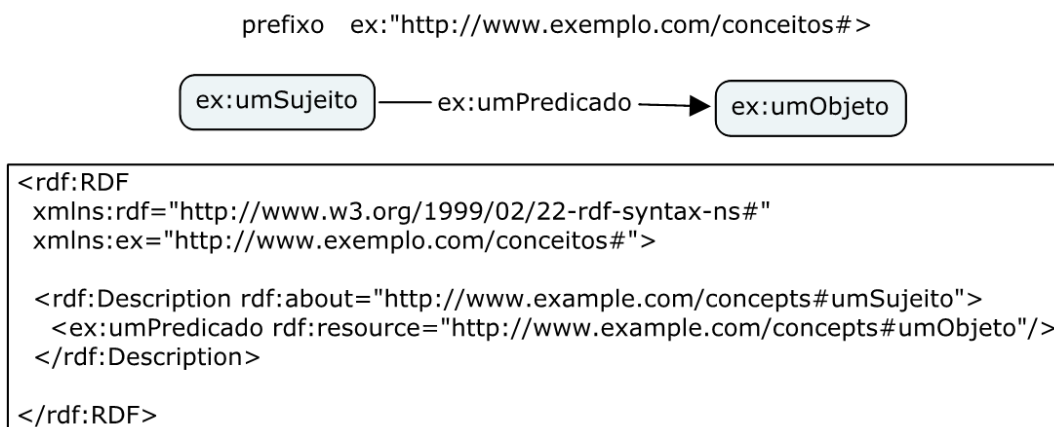
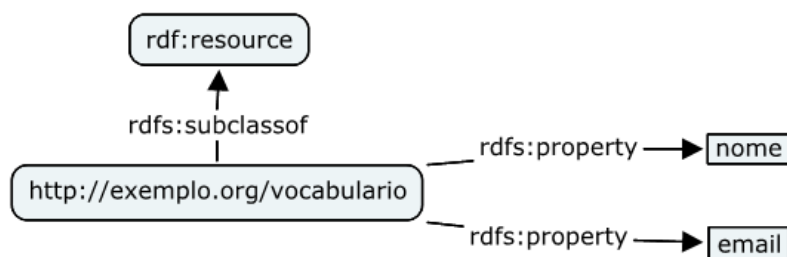


Figura 18 - Esquema gráfico e respectivo código de tripla RDF, conforme adaptação a partir de W3C (2005a).

A RDFS pode ser entendida como um meta-vocabulário, uma base para a construção de vocabulários de diferentes domínios e para diferentes propósitos. A figura 18 mostra um exemplo de definição de um vocabulário, uma especificação de domínio, como sendo um tipo de recurso (*rdf:resource*) com duas propriedades (“nome” e “email”).



```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

  <rdf:Description rdf:about="vocabulario">
    <rdfs:comment>Este meu vocabulario descreve um dominio de conhecimento
    </rdfs:comment>
  </rdf:Description>

  <rdfs:Class rdf:about="http://exemplo.org/vocabulario#Pessoa">
    <rdfs:isDefinedBy rdf:resource="http://exemplo.org/vocabulario#" />
    <rdfs:label>Pessoa</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource" />
  </rdfs:Class>

  <rdfs:Property rdf:about="http://exemplo.org/vocabulario#nome">
    <rdfs:isDefinedBy rdf:resource="http://exemplo.org/vocabulario#" />
  </rdfs:Property>

  <rdfs:Property rdf:about="http://exemplo.org/vocabulario#email">
    <rdfs:isDefinedBy rdf:resource="http://exemplo.org/vocabulario#" />
  </rdfs:Property>
</rdf:RDF>

```

Figura 19 - Exemplo de definição de vocabulário com RDF/RDFS.

Por outro lado, se houverem diversos vocabulários com definições dos mesmos conceitos, haverá ambiguidades. O importante é formar vocabulários bem definidos que sejam compartilhados como referências únicas para conceitos.

Conforme o exemplo da figura 20, ao invés de cada aplicação de domínio criar seu próprio vocabulário para os conceitos de “nome” e “email”, é possível utilizar o vocabulário aberto FOAF<sup>17</sup> (*Friend of a Friend*, Amigo de um Amigo) que possui as propriedades de semântica similar, “name” e “mbox”, respectivamente.

O Anexo C mostra alguns dos vocabulários abertos mais destacados, seus endereços na Internet e seus propósitos.

Vale ressaltar que, mesmo atendendo a domínios e propósitos diferentes, os vocabulários RDF/RDFS são componentes de uma arquitetura escalável e que

<sup>17</sup> Disponível em: <http://www.foaf-project.org/>. Acesso em: 3 fev. 2013.

prima pelo uso de padrões. Isto possibilita a combinação de conexões de triplas com diferentes vocabulários em diferentes níveis de abstração.

A figura 19 mostra a combinação de elementos do vocabulário FOAF (de menor abstração), com o elementos do vocabulário RDF (de maior abstração). É interessante observar que as duas estruturas que representam pessoas compartilham da mesma classe do tipo “foaf:person” do vocabulário controlado FOAF, o que garante unicidade de terminologia na representação. O Apêndice J mostra as triplas da representação da figura 19.

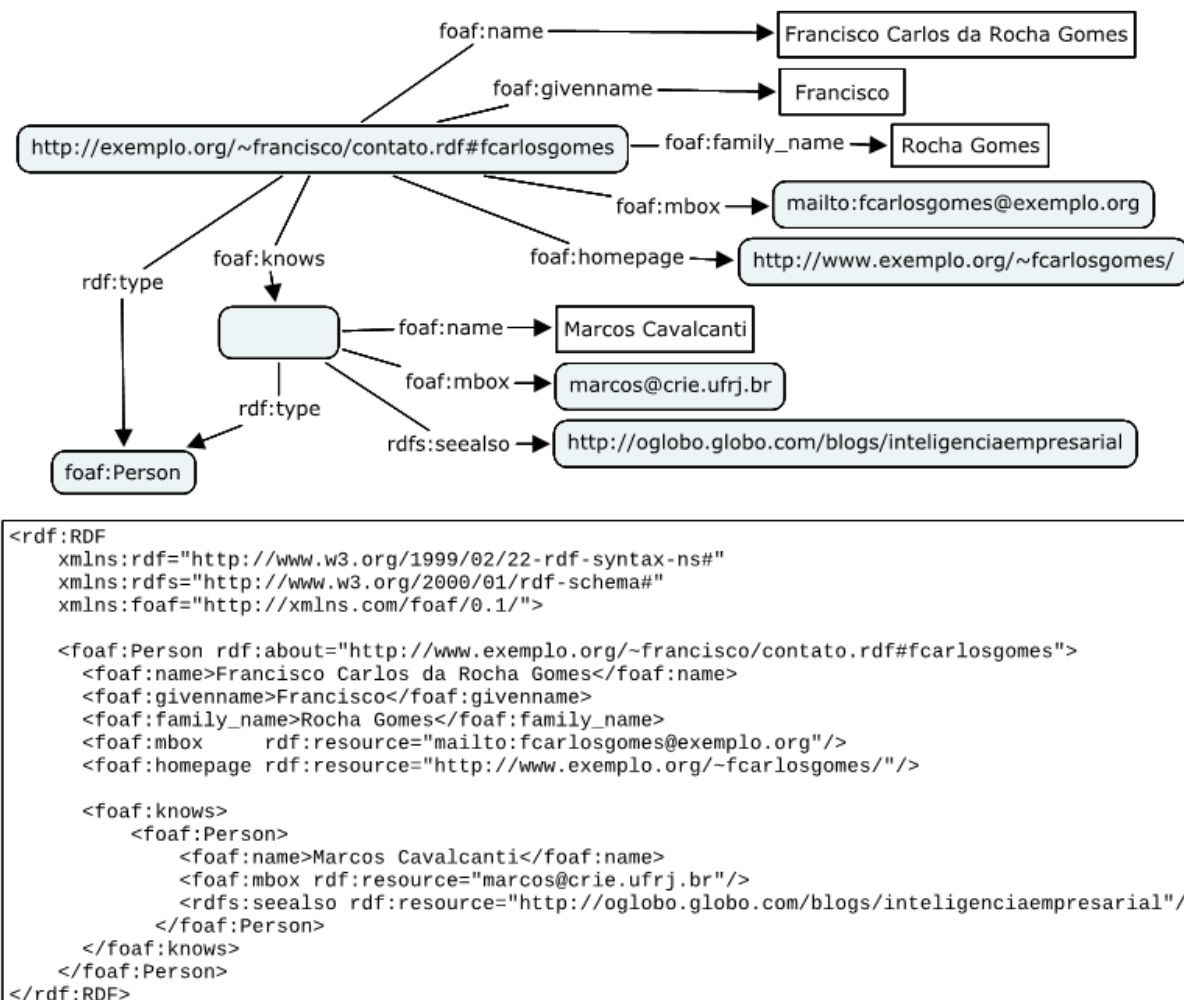


Figura 20 - Esquema gráfico e respectivo código de especificação com os vocabulários RDF e FOAF (*Friend of a Friend*, Amigo de um Amigo)

### 2.9.3 OWL – Linguagem para Construção de Ontologias na Web

A linguagem RDF/RDF-S é apropriada para ligações de dados e elaboração de vocabulários. Porém, seus construtos não possuem capacidade maior para expressar descrição lógica suficiente, afim de que, por meio de inferências automáticas, novas informações ou novas relações de dados possam ser geradas.

Antoniou e Harmelen (2008, p. 115-116) elencam as seguintes capacidades que não são possíveis por meio de RDF/RDFS:

- a) escopo local para propriedades de classes: por exemplo, em uma propriedade “come” não é possível expressar que vacas comem somente plantas e outros animais devam comer carne, também;
- b) disjunção de classes: às vezes é preciso dizer que duas classes (conceitos) são disjuntas como, por exemplo, “masculino” e “feminino”;
- c) combinações booleanas (lógica binária) de classes: às vezes é preciso construir classes que combinem outras classes por meio de relações como “união”, “intersecção”, “negação” e “complemento”. A classe “pessoa” pode ser uma união disjunta das classes “homem” e “mulher”;
- d) restrições de cardinalidade: em certos casos é preciso restringir o número de relações dos indivíduos de uma classe. Por exemplo, é preciso dizer que uma “pessoa” é nascida de um “pai” e uma “mãe” (que também são pessoas);
- e) atribuir características especiais às propriedades: em certas ocasiões é preciso dizer que uma relação é “transitiva” (como em “maior-do-que”), “única” como em “é\_mãe-de”), ou “inversa” de outra propriedade (como as propriedades “come” e “é-comido-por”).

Para permitir essas capacidades, linguagem OWL (*Ontology Web Language*, Linguagem de Ontologia da Web) agrega de forma escalável as características do RDF, as propriedades extendidas da RDF-S, juntamente com outros construtos desenvolvidos para criação de ontologias, aumentando a expressividade semântica da informação codificada.

Dentre as funcionalidades que a OWL pode realizar destaca-se: comparações (equivalência, diferença); operações de conjuntos como união, disjunção e intersecção; restrições de dados; transitividade, inversão de sentido da relação sujeito-predicado-objeto. (LACY, 2005, p. 167-212)

Conceitos, papéis e elementos individuais de um domínio de conhecimento podem ser representados, respectivamente, na forma de classes, propriedades e instâncias da linguagem OWL. (NEBOT; BERLANGA, 2012, p. 853)

Além disso, segundo Heath e Bizer (2011, p. 16-17), as características de conectividade da linguagem RDF, e de mediação semântica da RDF-S e OWL, permitem criar estruturas flexíveis de dados que podem contribuir para a ligação de dados heterogêneos acessados a partir diferentes fontes.

Porém, a linguagem OWL é muito genérica em procurar representar qualquer tipo conhecimento, tornando-se, assim, um tanto complicada. E, como o propósito desta linguagem é permitir o processamento das representações por meio de computadores, os arquitetos responsáveis por sua concepção a dividiram em três dialetos, cada uma com uma capacidade crescente em relação à anterior quanto ao processamento por parte de programas de raciocínio automático, chamados de *reasoners*.

Segundo Lacy (2005, p. 137-139) , estes três dialetos são:

- a) OWL Lite: versão mais simples da OWL, que possibilita uma execução mais eficiente por parte dos *reasoners*. É apropriada para a representação da informação em banco de dados;
- b) OWL DL: versão intermediária que possui elementos da OWL Lite, que agrega ainda construtos de lógica descritiva. Exige *reasoners* mais robustos;
- c) OWL Full: versão completa da linguagem, que possui bastante flexibilidade de expressão, mas exige alto grau de conhecimento de modelagem para aplicações específicas e implementação de *reasoners* mais complicados.

Por meio dos construtos da OWL é possível desenvolver um vocabulário para organização do conhecimento. No vocabulário SKOS (*Simple Knowledge*

*Organization System*, Sistema Simples de Organização do Conhecimento) alguns de seus elementos são construídos com OWL. Por exemplo:

- a) o termo que denota “Conceito”, “skos:Concept”, é uma instância da classe da OWL, “owl:Class”, ou seja, é derivado desta classe, mas não é considerada uma sub-classe<sup>18</sup>;
- b) as especificações para as relações semânticas “mais-especifico” e “mais-amplio”, “skos:narrower” e “skos:broader”, respectivamente, são construtos de vocabulário que mantêm uma relação inversa entre si. Isto é especificado por meio da OWL<sup>19</sup>.

No âmbito da organização do conhecimento, o vocabulário SKOS (*Simple Knowledge Organization System*, Sistema Simples de Organização do Conhecimento), destaca-se como um esquema recomendado pelo W3C para a criação de taxonomias, tesouros, esquemas de classificação, dentre outras estruturas (figura 21).

A grande vantagem do vocabulário SKOS é poder utilizar as capacidades do sistemas de organização do conhecimento (KOS) na arquitetura padronizada, escalável e integradora da Web Semântica. Os dados semânticos expressos em RDF podem estar integrados a sistemas de organização do conhecimento e ontologias para processamento de inferências e raciocínio automático, tudo sob a mesma arquitetura.

Segundo Yu (2011, p. 141-142), os sistemas simples de representação do conhecimento SKOS são formas de implementar organização e representação de conhecimento em computadores utilizando recursos de processamento semântico como a linguagem de descrição de recursos RDF, que pode ser utilizada para implementar vocabulários, taxonomias, tesouros, esquemas de classificação e listas de palavras-chave (*subject heading lists, tags*).

---

18 Disponível em: <http://www.w3.org/TR/skos-reference/#concepts>. Acesso em: 7 fev. 2013.

19 Disponível em: <http://www.w3.org/TR/skos-reference/#semantic-relations>. Acesso em: 7 fev. 2013.

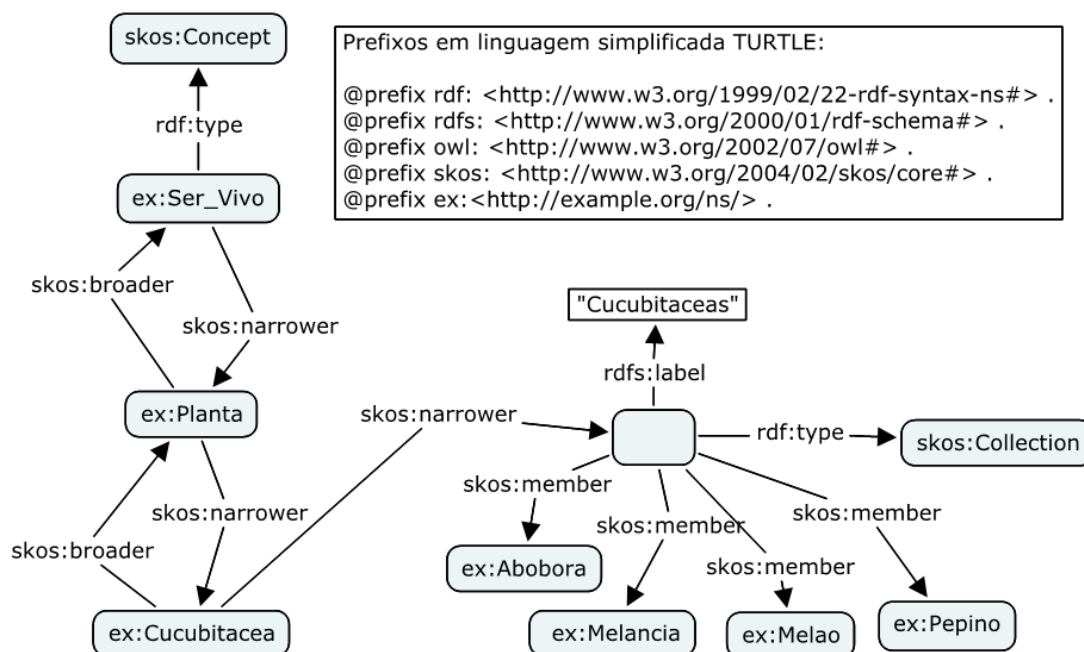


Figura 21 - Esquema de classificação utilizando SKOS.

## 2.9.4 Dados ligados (*Linked Data*)

Além das linguagens e vocabulários, outro componente importante do conhecimento explícito representado pelos recursos de Web Semântica são os dados e suas ligações (*Linked Data*, dados ligados), triplas que podem ser codificadas nos seguintes formatos:

- código RDF/XML: semelhante às especificações RDF de classes e propriedades, com extensas cadeias de caracteres;
- RDFa: codificação no formato RDF/XML capaz de ser inserida em um código de apresentação de dados HTML;
- Turtle: abreviação do formato RDF, suportando prefixos dos *namespaces* que referenciam os vocabulários;
- N-Triples: subconjunto do formato Turtle com menos recursos.

Segundo Berners-Lee (2006)<sup>20</sup>, não basta apenas colocar dados na Internet, é

<sup>20</sup> Mais detalhes em Yu (2011, p. 412-413).



preciso criar relações significativas entre eles e o uso de URIs é fundamental para o estabelecimento destas ligações. Para tanto, este autor desenvolveu quatro regras básicas para o governo dos dados ligados:

- a) as referências por URIs deve ligar coisas e não mais documentos e recursos como arquivos programas e dispositivos. “Coisas” passam a ser um nível de abstração maior, o que implica na maior amplitude e generalização do uso das ligações na Web;
- b) adoção de referências escritas em HTTP URI de maneira tornar legível o endereçamento pelas pessoas;
- c) quando houver alguma referência a dados, quando alguém procurar por uma URI, o serviço consultado deve ter condições de fornecer informações úteis, como no caso de uma consulta não retornar nenhum dado;
- d) deve ser incluindo outras referências a URIs de maneira que o usuário possa ter a capacidade de descobrir mais coisas.

De acordo com Heath e Bizer (2011, 69-80), a disponibilidade dos dados ligados pode ser feita da seguinte forma<sup>21</sup>:

- a) disponibilização de arquivos contendo triplas em RDF estáticos, sem atualização dos dados;
- b) conexão e conversão de banco de dados relacionais para o modelo de dados de triplas no formato RDF;
- c) conexão e conversão de outros tipos de arquivos de dados como os de formato texto, de planilha eletrônica, ou com delimitadores específicos.

Com base nas características dos dados ligados é possível deduzir uma arquitetura de integração de dados ligados, seguindo um mesmo padrão unificado, com padrões de acesso a dados auto-descritivos, possibilitando a descoberta de informações através do processamento de inferências nas hiperligações.

---

21 Disponível em: <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/#recipes>. Acesso em: 7 fev. 2013

### 3 ARQUITETURA DE REPOSITÓRIO SEMÂNTICO DE DADOS

Repositório é uma palavra originada do latim “*repositorium*”, que significa um “lugar onde se guarda, arquiva ou coleciona alguma coisa” (HOUAISS, 2009). Segundo Cunha e Cavalcanti, é um “lugar físico ou digital onde objetos são armazenados por um período de tempo, podendo ser consultados caso sejam solicitados pelos usuários, atendendo suas necessidades de informação” (2008, p. 321-322). Trata-se, portanto, de um conceito amplo.

Repositório de dados é um termo empregado no contexto digital para denotar um lugar onde dados são armazenados em sua forma elementar ou agregada, contidos em arquivos de dados estruturados, não-estruturados ou semi-estruturados; documentos digitais; centros de dados; bancos dados; e armazéns de dados constituídos pela integração de fontes de dados heterogêneas.

McInerney lembra que repositório de dados é uma forma de conhecimento explícito, e que deve ser robusto e flexível o suficiente para suportarem as atualizações frequentes, pois o conhecimento é dinâmico, em constante mudança e evolução (2002, p. 1016).

Os termos “memória” e “repositório” remetem ao significado da palavra “armazém”, lugar onde o conhecimento explícito pode ser guardado para posterior recuperação. Neste sentido, Elst e Abecker (2002, p. 362) defendem o estabelecimento de memórias organizacionais distribuídas, por ser muito caro sustentar modelos centralizados, onde a formulação de concordâncias entre as informações é mais difícil e rígida face às mudanças das necessidades informacionais. Em um modelo distribuído, o problema da integração de várias fontes de informação heterogêneas é abordado pela introdução de um nível uniforme de conhecimentos baseado em ontologias.

Inmon et al. (2008, p. 43, 48-49, 61-62, 81, 93) abordam a necessidade de se estabelecer um repositório corporativo de metadados de negócios e metadados técnicos. Os metadados de negócios descrevem elementos dos domínios de conhecimento do negócio obtidos a partir de descrições de processos e conteúdos

de documentos, enquanto que os metadados técnicos descrevem os elementos tecnológicos como nomes de tabelas e campos de bancos de dados, processos, estruturas, etc.

Segundo Kiryakov e Damova (2011, p. 233), repositórios semânticos são sistemas que combinam características dos sistemas de gestão de dados (SGBD) e motores de inferência, capazes de lidar com dados estruturados, levando em consideração sua semântica.

No âmbito corporativo, Dieng-Kuntz (2005, p. 67-80) propõe uma memória corporativa por meio de Web Semântica, denominada Web Semântica Corporativa, composta por uma tríade de elementos: recursos, ontologias e anotações.

Os recursos correspondem às fontes de conhecimento, como os documentos de vários formatos, pessoas (agentes de informação), serviços, software, ou programas. Dependendo da granularidade, os recursos podem ser um documento inteiro ou em parte, um banco de dados inteiro ou apenas um registro. Um recurso pode ser qualquer elemento identificável e que pode ser referenciado. Podem ser externos ou internos à organização.

As ontologias descrevem o vocabulário conceitual compartilhado pela empresa, e compreende os conceitos de alto nível (de integração), conceitos do domínio da aplicação, a relações entre conceitos, instâncias e axiomas envolvendo conceitos e as relações entre conceitos.

Anotações semânticas correspondem às referências sobre os recursos a fim de que estes possam ser recuperados. São exemplos de anotações os índices de documentos, descrições das habilidades de pessoas, ou as características de serviços, softwares, com base nas ontologias e com difusão na intranet ou na Web corporativa.

Além disso os agentes envolvidos com os processos informacionais da corporação podem ter os seguintes papéis:

- a) geradores de conhecimento registrado: autores de documentos, softwares, ou serviços que constituem os recursos de memória da organização;
- b) usuários finais: que fazem uso dos recursos;
- c) anotadores: geradores de conhecimento registrado ou mediadores que

produzem os descritores da informação registrada.

No âmbito de uma organização, uma arquitetura de dados ligados e integrados pode constituir um repositório semântico de dados<sup>22</sup>, componente essencial da memória organizacional, que possibilita o reuso do conhecimento explícito representado pelos dados semânticos, proporcionando a descoberta de novas informações a partir de padrões emergentes identificados por agregação e mineração de dados.

Arquitetura é a combinação, arranjo e organização de elementos e estruturas para formar um todo edificado emergente. Na perspectiva da complexidade, um resultado arquitetônico pode ser qualitativamente superior à soma dos elementos que o constituem.

No contexto da informação, arquitetura está muito relacionada com tecnologia, às vezes confundindo-se ambos os termos. Pois muitos dos produtos resultantes de uma concepção arquitetônica se apresenta como um artefato tecnológico. Por outro lado, são conceitos distintos. Inmon distingue o conceito de arquitetura do conceito de tecnologia:

Um aspecto confuso do armazenamento de dados é que ele é uma arquitetura, e não uma tecnologia. Isso frustra o técnico e o capitalista porque essas pessoas querem comprar algo numa caixa. Mas armazenamento de dados simplesmente não se presta a ser "espremido" numa caixa.

A diferença entre uma arquitetura e uma tecnologia é como a diferença entre Santa Fé, Novo México, e os tijolos de adobe.

Se você dirigir pelas ruas de Santa Fé, você sabe que está lá e em nenhum outro lugar. Cada casa, cada prédio, cada restaurante tem uma aparência distinta, que diz: "Esta é Santa Fé." O visual e o estilo que fazem de Santa Fé algo distinto é a sua arquitetura. Por outro lado, a arquitetura é composta de coisas como tijolos e vigas expostas.

Há uma arte inteira na confecção de tijolos e vigas expostas. E é verdade que você não poderia ter de Santa Fe arquitetura sem ter tijolos e vigas expostas.

Mas tijolos e vigas expostas por si só não fazem uma arquitetura.

Esses componentes são tecnologias independentes. Por exemplo, você tem tijolos de adobe em todo o Sudoeste e no resto do mundo que não são semelhantes aos da arquitetura de Santa Fé.

Assim é com a arquitetura de armazenamento de dados, de bases de dados, e sua tecnologia subjacente, que não é a arquitetura. São duas coisas muito diferentes.

---

<sup>22</sup> Detalhes sobre arquitetura de repositório semântico para gestão do conhecimento pode ser visto em Kiryakov et al. (2009, p. 71)

É inquestionável que existe uma relação entre armazenamento de dados, base de dados e tecnologia, mas certamente não é a mesma coisa. O que existe é que o armazenamento de dados requer o apoio de muitos tipos diferentes de tecnologia. (INMON, p. 2002, p. xv)

A figura 22 mostra uma arquitetura de repositório semântico de dados, proposta por Berners-lee (2005), onde destacam-se os dados legados, dados semânticos padronizados e integrados, e a camada superior de aplicações de acesso aos dados<sup>23</sup>.

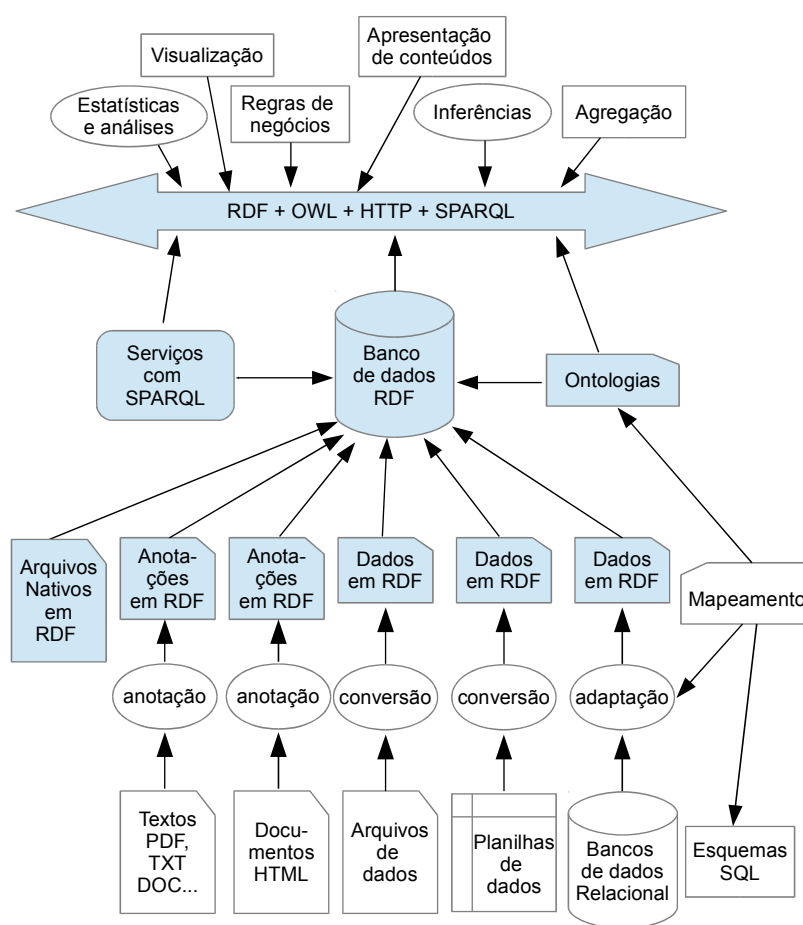


Figura 22 - Arquitetura de Integração de dados ligados em RDF a partir de fontes heterogêneas (adaptado de BERNERS-LEE, 2005)

<sup>23</sup> Uma visão para implementação deste modelo pode ser vista em Hebler et al. (2009, p. 468; 474).

A camada de dados legados é constituída por diversas fontes de dados de diferentes formatos e bancos de dados, onde subconjunto ou registros de bancos de dados, arquivos de dados (estruturados, semi-estruturados e não estruturados) e planilhas, podem ser mapeados para triplas RDF por meio de programas mediadores. Além disso, documentos e arquivos de texto, imagens, som, etc, podem ser ligados por meio de anotações de metadados por meio de código RDFa em documentos HTML ou em arquivos RDF (figura 22), e classificados por mecanismos de organização do conhecimento sob a forma de triplas RDF (SKOS).

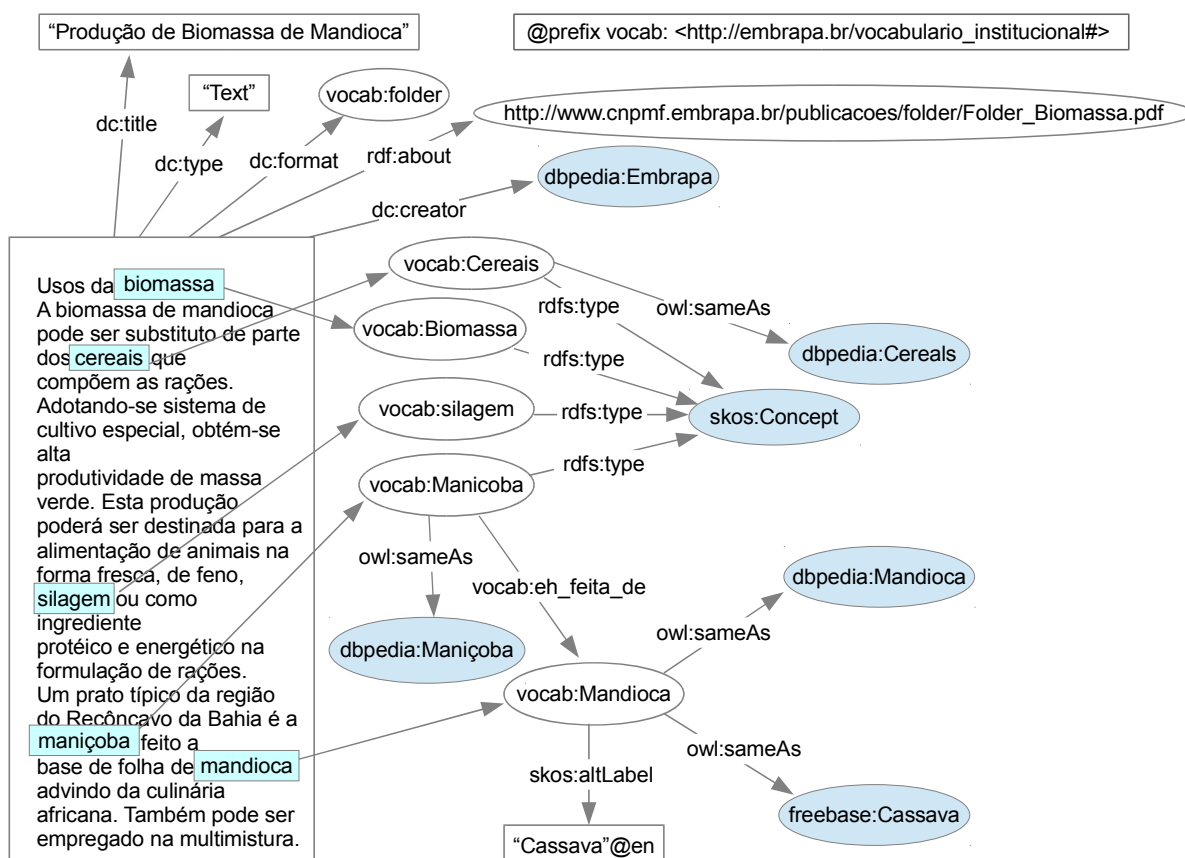


Figura 23 - Esquema conceitual de anotação de trecho de documento tipo texto<sup>24</sup>.

A camada semântica, que compreende três elementos: repositório semântico de dados integrados; ontologias (Tbox), vocabulários e outros esquemas de lógica; e

24 Detalhes sobre os termos da DBPedia utilizados neste esquema podem ser vistos no Anexo E.

serviços de protocolo SPARQL para consultas aos dados.

Os vocabulários asseguram a unicidade da informação evitando as ambiguidades e redundâncias, proporcionando um ambiente de compartilhamento e interoperabilidade.

A camada de dados semânticos provê um conjunto de serviços de dados baseados no padrão Web (RDF, OWL, HTTP, SPARQL). A partir desse conjunto de serviços diversas aplicações podem ser desenvolvidas como, por exemplo, definição de regras de negócios, análises e estatísticas, inferências, apresentação de conteúdos e visualização de estruturas de dados. Além disso, sistemas de gestão de conteúdos (CMS - *Content Management Systems*) podem estar conectados à camada de dados semânticos<sup>25</sup>.

Ferramentas de gestão do conhecimento, especificamente para análise e mineração de dados podem operar diretamente sobre os dados em triplas, ou indiretamente por meio das ontologias e recursos de organização do conhecimento como (vocabulários, taxonomias, tesouros, classificações facetadas, etc).

Quanto ao processo de desenvolvimento da arquitetura Dieng-Kuntz (*op. cit*) delinea um processo de oito etapas de construção da Web Semântica Corporativa:

- 1) inventário das instalações (diagnóstico e contextualização);
- 2) escolha de cenário de aplicação: definição dos objetivos dos usuários e das aplicações; identificação das fontes de informação; identificação de conteúdos e de partes de ontologia;
- 3) construção de ontologia a partir das pessoas (conhecimento tácito), dos *corpus* textuais (conteúdo documentado) e dos acervos de informações estabelecidas (banco de dados, arquivos);
- 4) validação das ontologias: consistência e checagem, auditoria e validação, avaliação pelos usuários finais;
- 5) constituição, organização e validação dos novos recursos criados ou dos recursos legados;
- 6) anotação dos recursos: de forma manual, automática (por meio de

---

25 Exemplos de conexão entre CMS e dados ligados em RDF com Drupal:  
<http://drupal.org/project/rdfx> e  
<http://www.ibm.com/developerworks/library/wa-datasets/>. Acesso em: 8 fev. 2013.

*softwares*), ou de forma semiautomática (mista);

- 7) validação das anotações e da Web Semântica corporativa: consistência e checagem, auditoria e validação, avaliação pelos usuários finais;
- 8) manutenção e acompanhamento evolutivo dos componentes da Web Semântica Corporativa: ontologias, recursos e anotações.

Por outro lado, especificamente quanto à metodologia de construção de ontologias, destaca-se, dentre outros, o método *Methontology* de Gómez-Pérez et al. (2004, p. 125), um conjunto de atividades divididas em cinco fases que compreendem o ciclo de vida da ontologia. Estas fases são as seguintes: especificação, conceitualização, formalização, implementação e manutenção (figura 24).

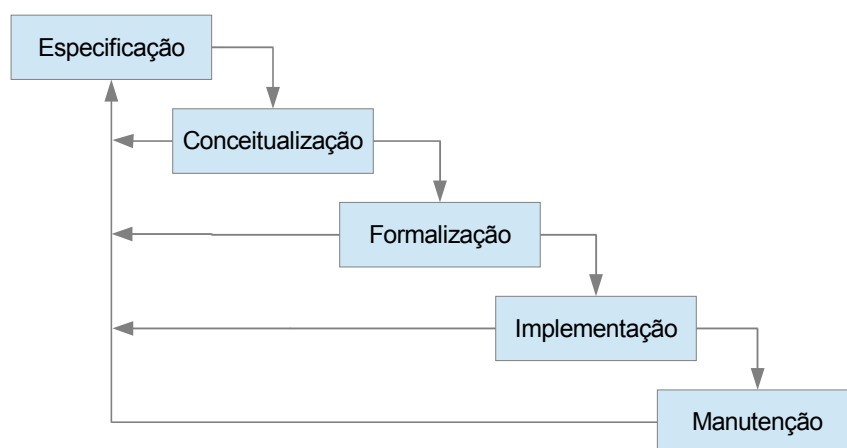


Figura 24 - Atividades da metodologia para construção de ontologias Methontology (GÓMEZ-PÉREZ et al., 2004, p. 127).

Segundo Breitman (2010, p. 73), na fase de especificação questionamentos iniciais são levantados sobre o escopo e os objetivos da ontologia, bem como a identificação de seus usuários.

A fase de conceitualização compreende as atividades da seguinte ordem:

- 1) construção de glossário de termos;
- 2) construção de taxonomias de conceitos;
- 3) construção provisória de relações binárias entre conceitos;
- 4) construção de dicionário de conceitos;



- 5) descrição provisória de relações binárias;
- 6) descrição de atributos de instâncias;
- 7) descrição de atributos de classes;
- 8) descrição de constantes;
- 9) descrição de axiomas formais;
- 10) descrição de regras;
- 11) descrição de instâncias.

A formalização compreende adequar o modelo conceitual para uma lógica formal (modelo lógico), independente de linguagem de implementação para computador.

A implementação implica na codificação da ontologia permitindo sua execução por computador. Nesta fase a ontologia é escrita nas linguagens da Web Semântica (RDF/RDFs, RDFa, OWL, etc.)

A manutenção consiste em corrigir problemas e adequar as ontologias a novos requisitos, conduzindo para uma melhoria contínua das ontologias, num processo cíclico.

Em síntese, uma arquitetura de repositório de dados baseado em Web Semântica permite a unicidade, padronização e escalabilidade da informação em âmbito corporativo. A unicidade e padronização da informação promove uma camada de acesso uniforme para as aplicações que a utilizam. Já a adoção de padrões universais promove a interoperabilidade entre sistemas ou bases de conhecimento de diferentes instituições. Finalmente, a escalabilidade promove a agregação de novos elementos ao repositório sem nenhum ou pouco custo de modificação nas estruturas e especificações já estabelecidas.

No contexto atual em que as instituições de pesquisa necessitam cada vez mais de recursos que auxiliem na produção do conhecimento científico, o ideal é que os repositórios de dados tenham um papel ativo no processo de produção de conhecimento, onde novos conhecimentos possam emergir a partir da complexa combinação das informações integradas e armazenadas.

Particularmente no caso dos repositórios de dados científicos, os

pesquisadores são confrontados com problemas de gestão de dados significativos, devido ao grande volume e alta complexidade dos dados, sendo necessário usar mecanismos de mediação integrativa e semântica que possa resolver a abrangência e a heterogeneidade dos dados (LUDASCHER et al., 2006, p. 109).

### 3.1 Arquitetura de repositório semântico

Com o objetivo de caracterizar a arquitetura de repositório semântico em um domínio específico de aplicação, este trabalho apresenta uma proposta de arquitetura para instituição de pesquisa agropecuária, especificamente para a Empresa Brasileira de Pesquisa Agropecuária – Embrapa.

De acordo com o seu portal corporativo<sup>26</sup>, a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) é uma organização vinculada ao Ministério da Agricultura, Pecuária e Abastecimento, criada em 26 de abril de 1973, cuja missão é “viabilizar soluções de pesquisa, desenvolvimento e inovação para a sustentabilidade da agricultura, em benefício da sociedade brasileira.”

Sua atuação se dá em quase todos os estados brasileiros (Anexo H) nos mais diferentes biomas do país, além de atuar no exterior por meio de parcerias com laboratórios nos Estados Unidos e na Europa (França, Alemanha e Inglaterra).

Com objetivo de disseminar tecnologias e inovações da agricultura tropical para o desenvolvimento agrícola de países em desenvolvimento, a Embrapa atua em Gana (Embrapa África), Venezuela, no Panamá (Embrapa Américas, com atuação na América Central e Caribe).

A Embrapa é um sistema formado por Unidades Administrativas, também denominadas Unidades Centrais, localizadas no edifício-sede em Brasília/DF, e por Unidades de Pesquisa e de Serviços, também chamadas Unidades Descentralizadas, distribuídas nas diversas regiões do Brasil (Anexo G).

Suas Unidades Descentralizadas são assim classificadas<sup>27</sup> como:

a) Unidades de serviço;

---

26 Disponível em: [http://www.embrapa.br/a\\_embrapa/missao\\_e\\_atuacao](http://www.embrapa.br/a_embrapa/missao_e_atuacao). Acesso em: 10 fev. 2013.

27 Disponível em: [http://www.embrapa.br/a\\_embrapa/unidades-de-pesquisa-e-de-servicos/unidades-de-pesquisa](http://www.embrapa.br/a_embrapa/unidades-de-pesquisa-e-de-servicos/unidades-de-pesquisa). Acesso em: 10 fev. 2013.

- b) Unidades de pesquisa de produtos;
- c) Unidades de pesquisa de temas básicos;
- d) Unidades de pesquisa ecorregionais.

O Centro de Pesquisa Agroflorestal do Acre, Embrapa Acre, é uma Unidade de pesquisa ecorregional localizada no Estado do Acre, região ocidental do Brasil, fronteira entre áreas de produção agrícola e biodiversidades intocadas da Amazônia brasileira<sup>28</sup>.

Segundo seu Regimento Interno<sup>29</sup> as finalidades da Embrapa Acre, dentre outras, são:

- a) atuação com ênfase no Estado do Acre;
- b) apoiar Sistema Nacional de Pesquisa Agropecuária – SNPA;
- c) fazer pesquisa, desenvolvimento e inovação de produtos e serviços demandados pelo agronegócio;
- d) viabilizar o desenvolvimento de produtos e serviços dos setores: agrícola, pecuário, florestal e agroindustrial;
- e) promover e viabilizar a caracterização e o zoneamento de recursos naturais.

A Embrapa Acre atua em quatro núcleos de pesquisa<sup>30</sup>:

- a) Fruticultura e Plantas Nativas Agroindustriais;
- b) Produção Animal Sustentável;
- c) Manejo Florestal Sustentável e Silvicultura;
- d) Sistemas Integrados e Recuperação de Áreas Degradadas.

A partir de entendimentos dos conceitos sobre os quatro núcleos de pesquisa, bem como da análise de conteúdo do portal de Internet, das publicações, e especialidades dos pesquisadores da Embrapa Acre<sup>31</sup>, foi possível decodificar as

28 Disponível em: [http://www.cpfac.embrapa.br/a\\_unidade/localizacao](http://www.cpfac.embrapa.br/a_unidade/localizacao). Acesso em: 10 fev. 2013.

29 Disponível em: <http://hotsites.sct.embrapa.br/acessoainformacao/institucional/regimentos>. Acesso em: 10 fev. 2013.

30 Disponível em: [http://www.cpfac.embrapa.br/a\\_unidade/atribuicoes/](http://www.cpfac.embrapa.br/a_unidade/atribuicoes/). Acesso em: 10 fev. 2013.

31 Disponível em: <http://www.cpfac.embrapa.br/publicacoes>. e [http://www.cpfac.embrapa.br/a\\_unidade/equipe](http://www.cpfac.embrapa.br/a_unidade/equipe). Acesso em: 10. fev. 2013.

áreas de conhecimento e de pesquisa.

As principais áreas do conhecimento em que atua a Embrapa Acre são: agricultura, pecuária, floresta, agroindústria e socioeconomia. Os detalhes destas áreas estão nos Apêndices L.

As principais áreas de pesquisa são: estudos de solos, produção vegetal , produção animal, floresta, pós-colheita, socioeconomia. Os detalhes destas áreas estão no apêndice M.

Ainda de acordo com a análise de conteúdo das páginas na Internet, os principais agentes que produzem, difundem e usam informação são:

a) clientes internos<sup>32</sup>:

- pesquisadores;
- analistas;
- assistentes;

d) clientes externos:

- produtores;
- governo;
- empresas da iniciativa privada;
- instituições de pesquisa nacionais e internacionais;
- outras Unidades da Embrapa no âmbito nacional.

Em termos de estrutura orgânica, funcional a Embrapa Acre pode ser classificada da seguinte forma:

a) área de pesquisa, com:

- laboratórios<sup>33</sup>:

- bromatologia;
- solos;
- morfogênese e Biologia Molecular;
- fitopatologia;
- tecnologia de Alimentos;
- óleos essenciais;

---

32 Disponível em: [http://www.cpfac.embrapa.br/a\\_unidade/equipe](http://www.cpfac.embrapa.br/a_unidade/equipe). Acesso em: 10. fev. 2013.

33 Disponível em: <http://www.cpfac.embrapa.br/laboratorios>. Acesso em: 10 fev. 2013.

- entomologia;
- farinha;
- campo experimental;
- b) transferência de tecnologia;
- c) administração de apoio.

### **3.1.1 Organização e representação do conhecimento da arquitetura**

A partir da análise dos conteúdos dos portais da Embrapa Sede, das Unidades descentralizadas, particularmente da Embrapa Acre, foi constatada uma boa organização e padronização de conteúdos e processos que facilitam a interação do usuário com o conhecimento explícito ali armazenado. As informações estão bem estruturadas em taxonomias numa disposição gráfica uniforme. A facilidade de acesso é potencializada por mecanismos de busca por meio de palavras-chave, que permitem a rápida localização e recuperação de conteúdos de páginas da Web, dados, imagens e outros documentos.

Neste caso, a gestão do conhecimento é tipicamente voltada para gestão de conteúdo e gestão de documentos, com processos bem definidos. É uma situação adequada para a interação do usuário.

Porém, considerando o crescente volume de documentos e das relações entre eles, surge a necessidade de uma maior intervenção e apoio computacional sobre as informações armazenadas.

Do ponto de vista do usuário, o conhecimento explícito disponível nos portais é completo e íntegro. Mas, do ponto de vista do processamento dos computadores, a granularidade das informações está, em sua maioria, no nível de documento. Na perspectiva do processamento semântico, dos dados ligados, do processamento de inferências e integração de dados, a informação encontra-se fragmentada e ambígua, o que dificulta a recuperação do conhecimento emergente.

Neste sentido, destacam-se algumas observações levantadas na análise de conteúdo dos portais:

- a) no portal da Embrapa Acre existem áreas de conhecimento e de pesquisa que compartilham da designação pelo mesmo termo, porém, com diferentes estruturas e relações <sup>34</sup>. É o caso, por exemplo, dos termos “Produção Animal” e “Floresta” que possuem detalhamentos estruturais diferenciados, e o termo “entomologia” que designa tanto um laboratório quanto uma área do conhecimento. Provavelmente os mesmos termos são utilizados em contextos diferentes mas que possuem alguma relação entre si, mas esta análise de conteúdo somente é possível manualmente por meio do acesso a cada conteúdo dos documentos. Não há uma forma de ligar os termos ou detalhá-los para recuperar essa perspectiva de termos semelhantes em contextos diferenciados;
- b) não há uma forma de recuperação dos detalhamentos estruturais e referenciais por algum mecanismo integrador a partir os mesmos dados que compõem os documentos;
- c) por falta de definição ou aplicação de um vocabulário corporativo implementado em computador, há ambiguidades de termos<sup>35</sup> e não há como recuperar diferentes combinações de informações.

Portanto, a construção de uma arquitetura de repositório semântico é fundamental para a gestão do conhecimento, geração de conhecimento emergente e inovação, num contexto de grandes volumes de dados, diferentes necessidades dos usuários, diferentes perspectivas de percepção do conhecimento, em diferentes momentos e circunstâncias

---

34 Ver apêndice G sobre as áreas de conhecimento e áreas de pesquisa da Embrapa Acre.

35 Por exemplo, ambiguidades dos nomes das titularidades e das áreas do conhecimento dos pesquisadores integrantes da equipe a da Embrapa Acre disponível em: [http://www.cpafac.embrapa.br/a\\_unidade/equipe](http://www.cpafac.embrapa.br/a_unidade/equipe). Acesso em: 11 fev. 2013. Este é um exemplo de ambiguidade no conteúdo do portal de uma Unidade da Embrapa, comparando-se os conteúdos de outras Unidades se verifica também novas ambiguidades.

### 3.1.2 A arquitetura

Com base nas metodologias expostas, este trabalho propõe uma Arquitetura de repositório semântico de Unidade de Pesquisa da Embrapa.

O esquema da figura 25 representa a arquitetura. Os círculos representam entidades que são fontes de dados ligados e ontologias. O círculo maior mostra os componentes internos que detalham a funcionalidade da arquitetura em uma unidade de pesquisa da Embrapa. Dois círculos menores à direita e mais abaixo do esquema representam outras abstrações de unidades da Embrapa: Uma unidade de Pesquisa e a Embrapa Sede. Por sua vez, o círculo médio mais ao alto à direita representa o conjunto de dados ligados e ontologias disponíveis por outras entidades fora do âmbito da Embrapa.

As setas indicam o sentido do fluxo de informações. O sentido de saída de um componente da arquitetura indica que a informação é fornecida ou registrada em outro componente. O sentido de entrada indica a leitura de informações a partir de um componente. Setas duplas indicam o duplo sentido do fluxo tanto de registro quanto de leitura de informações.

O círculo maior apresenta em detalhes da arquitetura. O repositório semântico ocupa a parte central e é constituído pela integração de fontes heterogêneas de dados ligados em RDF (tripas), proporcionando uma interface de dados uniforme, um barramento de informação (*information bus*) para as aplicações que se conectam com o repositório. Esta integração de dados pode ser virtual, como uma federação de dados (*data federation*) sem a necessidade da constituição de um repositório físico.

O repositório de dados é formado por dados ligados e ontologias. Os dados são integrados a partir das seguintes fontes:

a) dados legados originados:

- da conversão de arquivos de dados em diferentes formatos (csv, txt, etc);
- da conversão de planilhas de dados;
- da anotação ou indexação de metadados de documentos como páginas Web (HTML), arquivos textos (doc, pdf, txt), planilhas eletrônicas (xls),

- arquivos de dados não-estruturados;
- triplificação<sup>36</sup> (conversão) de bancos de dados relacionais para RDF;
- b) dados nativos em RDF localizados em arquivos estáticos, ou gerados dinamicamente por aplicações.

Dentre as ontologias que podem ser armazenadas no repositório destacam-se:

- a) as ontologias de domínio que formam o vocabulário controlado corporativo da Unidade que garante a unicidade (não-ambiguidade) da informação;
- b) outras ontologias de representação, com axiomas, regras de negócio.

A arquitetura mantém uma interface com o meio externo disponibilizando dados ligados e ontologias abertos para outras entidades como unidades de pesquisa da Embrapa, Embrapa Sede, além de poder compor os dados integrados do programa de Governo eletrônico (e-GOV) e outros repositórios externos.

Por outro lado, as entidades externas geradoras de dados ligados e ontologias também podem disponibilizar seus dados, que podem ser integrados ao repositório semântico em questão. Dentre os repositórios externos destaca-se o AGROVOC da FAO<sup>37</sup> (Anexo I).

Na camada de aplicações destaca-se os aplicativos para visualização de dados semânticos, na qual pode-se observar as informações subjacentes àquelas definidas como foco inicial da consulta do usuário. Esta visão facilita a identificação das relações entre os dados ou, de forma agrupada, as relações entre conceitos. Além disso, diversas perspectivas, níveis de abstrações, e tipos de agregações, podem ser obtidas a partir de diferentes arranjos dos conteúdos semânticos. Isto pode contribuir para a emergência de novos padrões de dados ou novas percepções sobre os dados. É possível também obter visões hierárquicas ou em rede dos dados, ou ainda dispostos em uma árvore hiperbólica<sup>38</sup>.

36 Para mais detalhes acessar os conteúdos disponíveis em: <http://semanticweb.org/wiki/Triplify> e <http://semanticweb.org/wiki/Triplify>. Acesso em 2 de fev. de 2013.

37 Mais detalhes em: Caracciolo et al. The AGROVOC Linked Dataset, 2012. Disponível em: [http://www.semantic-web-journal.net/sites/default/files/AGROVOC\\_Dataset\\_vFinal\\_v2.4.pdf](http://www.semantic-web-journal.net/sites/default/files/AGROVOC_Dataset_vFinal_v2.4.pdf). Acesso em: 13 fev. 2013.

38 Uma lista de aplicativos de visualização de dados RDF pode ser encontrada em:



A adaptação dos bancos de dados legados para o formato de triplas pode ser realizada pelo processo de triplificação. É um processo de mapeamento no qual, por meio de regras, os dados relacionais podem ser representados na forma de triplas. A grande vantagem desse processo é de, por um lado, manter todas as aplicações de bases de dados relacionais operando e atualizando seus dados e, por outro lado, ter a capacidade de poder integrar esses dados com outros dados ligados<sup>39</sup>.

Documentos como textos, publicações em PDF, arquivos de multimídia, planilhas de dados, etc, podem ser marcados com metadados pelo processo de anotação, que pode ser manual, semiautomático ou automático. De forma manual as pessoas fazem a anotação diretamente no documento, quando este permite metadados, ou em arquivo subjacente de metadados. O processo de anotação semiautomático envolve o trabalho do usuário sobre os documentos por meio de softwares apropriados para tal finalidade. E o processo de anotação automática é realizado por aplicativos com códigos internos que geram anotações.

No processo de anotação cabe destacar que os metadados e os termos-chave (*tags*) devem estar em concordância com os vocabulários controlados vigentes da corporação.

Dadas estas características, a arquitetura pode criar condições mais favoráveis para que o conhecimento emergente possa ser obtido pela organização, por meio da integração e combinação de informações semânticas, e o reuso do conhecimento explícito registrado.

---

Large-scale RDF Graph Visualization Tools:  
<http://www.mkbergman.com/414/large-scale-rdf-graph-visualization-tools/>. Acesso em: 13 fev. 2013.

39 Sobre triplificação ver os links (acesso em: 13 fev. 2013):

Triplify expose semantics:

<http://triplify.org/Overview>

Tutorial de Triplificação da PUC-Rio:

<http://www.inf.puc-rio.br/~psalas/tutorial.html>

D2RQ - Accessing Relational Databases as Virtual RDF Graphs:

<http://d2rq.org/> e

<http://d2rq.org/d2r-server>

Mapping Relational Data to RDF with Virtuoso's RDF Views:

<http://virtuoso.openlinksw.com/whitepapers/relational%20rdf%20views%20mapping.html>

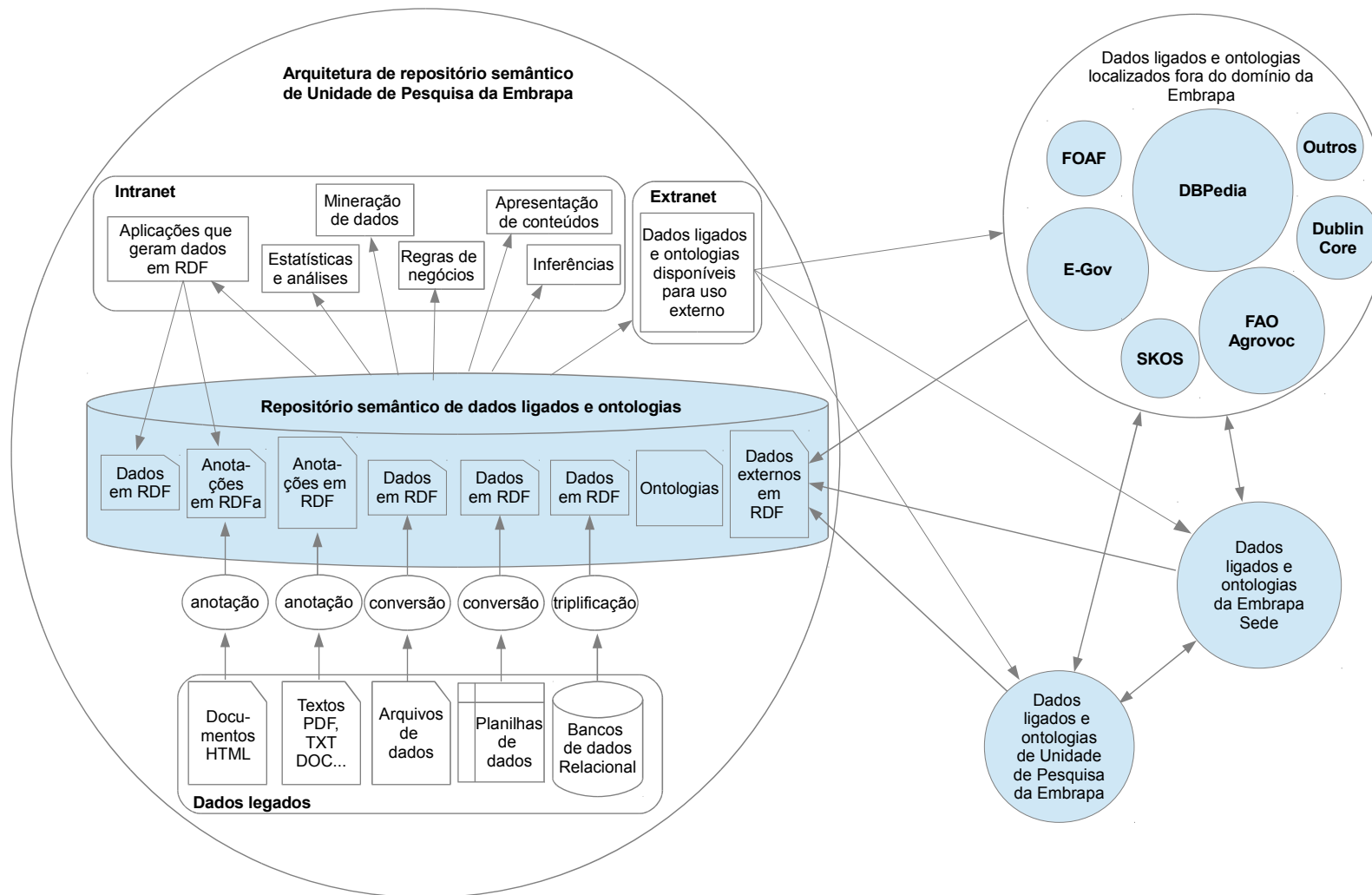


Figura 25 - Arquitetura de repositório semântico para domínio de aplicação da Embrapa.

### 3.1.3 Processo de desenvolvimento e manutenção da arquitetura

Tão importante quanto a concepção da arquitetura é a definição do processo de seu desenvolvimento e manutenção.

É fundamental para a implantação da arquitetura em ambiente corporativo a especificação do processo, suas atividades e os papéis dos agentes envolvidos, compreendendo, dentre outras, as seguintes atividades:

- a) definir política (governança) para gestão e controle de dados e metadados, particularmente quanto à sua organização, armazenamento e acesso;
- b) estabelecer comitê gestor para validar e manter a qualidade dos dados e metadados, bem como tratar da verificação da validade e confiabilidade das fontes de informação;
- c) especificar, implantar e promover o uso de um glossário corporativo com dicionário de dados sob a forma de vocabulário controlado para garantir unicidade da informação e seu acesso uniforme por toda corporação<sup>40</sup>. No contexto da Embrapa, o vocabulário pode ser dividido em dois níveis de gestão: um vocabulário central baseado na Unidade Sede da instituição; e vocabulários gerenciados em cada Unidade descentralizada, porém integrados e em concordância com o vocabulário central, sem ambiguidades ou conflitos entre termos;
- c) delinear e compartilhar modelos de domínio e subdomínios de conhecimento;
- d) criar repositório corporativo de metadados de negócios e técnicos, onde:
  - metadados de negócios descrevem os elementos dos domínios de conhecimento do negócio e que são obtidos a partir de descrições de processos e conteúdos de documentos;
  - metadados técnicos descrevem os elementos de artefatos tecnológicos como, por exemplo, nomes de tabelas e campos de bancos de dados;
- e) definir processo de estruturação física ou virtual de dados não-

---

40 Para mais detalhes ver Inmon et al., 2008, p. 61-62; 81; 93.

estruturados<sup>41</sup>;

- f) definir processo de padronização de dados, sua captura e política de uso;
- g) implementar representação de dados ligados por meio da linguagem RDF;
- h) implementar representação de domínios por meio da linguagem OWL;
- i) especificar e implementar ontologias;
- f) especificar e implementar restrições de segurança e confidencialidade;
- g) especificar e implementar processos de conversão de dados legados para RDF.

---

41 Para mais detalhes ver Inmon e Nesavich, 2008, p. 30.

## 4 CONCLUSÃO

### 4.1 Considerações finais

O conhecimento é muito mais útil quando agregado a outros conhecimentos e compartilhado com outras pessoas, mas, diante da grande complexidade do mundo atual, como se pode extrair conhecimento significativo de uma massa de informação cada vez maior?

Este trabalho mostrou a necessidade de representar melhor o conhecimento diante da demanda para se compreender fenômenos cada vez mais complexos, inter-relacionados direta ou indiretamente.

O desafio da representação da organização e representação do conhecimento é atender o usuário com a informação certa no contexto da natureza dinâmica do conhecimento, das múltiplas perspectivas de se olhar os fenômenos, dos contextos que mudam, e das possíveis combinações de informações.

É importante disponibilizar o acesso às diferentes combinações de informações, às informações subjacentes, correlatas, denotadas, inferidas, enfim, todas as informações que não são diretas ou óbvias, que contribuem para um nível de percepção superior da qual surge o novo conhecimento e a condição para a inovação.

É preciso saber lidar bem com as transições do conhecimento que está na mente das pessoas ou grupo (tácito) para o conhecimento registrado (explícito), mais reduzido em conteúdo e mais rígido, porém, menos complexo e necessário para o compartilhamento.

No âmbito corporativo o fluxo de conhecimento não depende apenas dos canais disponíveis e das condições de infraestrutura, mas também da forma como o conhecimento é armazenado e compartilhado, e isto depende da sua representação e organização. Pois é possível ter na corporação toda uma infraestrutura com meios de comunicação eficientes, porém o conhecimento será pouco utilizado se não estiver disponível numa arquitetura semântica organizada e flexível, e que facilite seu acesso e assimilação.

A memória de uma corporação pode ser como uma grande biblioteca, com milhares de livros. Se o usuário não tiver condições de saber o que pode obter das fontes, receberá apenas aquilo que seu entendimento acha que pode obter.

Neste sentido, uma interface que se limita a disponibilizar uma mera lista de itens sobre o que foi solicitado é mais limitada em possibilidades de mostrar o conteúdo da fonte do que a apresentação de uma rede semântica com os conceitos e suas relações subjacentes como uma árvore do conhecimento que mostre o todo interligado<sup>42</sup>.

Mas para se obter esta rede semântica é preciso que o conhecimento esteja contido numa arquitetura adequada. A arquitetura de repositório semântico permite que mecanismos de inferência automáticos operem sobre as ontologias e dados semânticos ligados de maneira que possa oferecer informações subjacentes às solicitadas pelos usuários. Estes autômatos poderão entender o que o usuário busca para encontrar o que ele precisa<sup>43</sup>.

Este trabalho procurou apresentar a problemática das corporações sobre a necessidade de converter o conhecimento explícito atual, baseados em documentos, para o nível de dados semânticos ligados, armazenados em fontes heterogêneas e integrados como um repositório de acesso centralizado e padronizado.

A arquitetura de repositório de dados semânticos representa uma oportunidade para a gestão do conhecimento criar um ambiente corporativo favorável para geração de novos conhecimentos emergentes, a partir do reuso do conhecimento explícito acumulado na memória da organização, com apoio da tecnologia para lidar com grandes volumes de dados e complexidade dos problemas.

Vale ressaltar que a arquitetura em questão não procura substituir o atual paradigma das bases de conhecimento relacionais e das bases de documentos, mas servir de complemento a estas arquiteturas tradicionais de informação. A arquitetura de repositório semântico não substitui as bases de dados científicas, mas constitui

---

42 Maturana, Humberto R.; Varela, Francisco J. *A árvore do conhecimento as bases biológicas da compreensão humana*. São Paulo. Editora Palas Athena, 2001. 288p.

43 Esta é uma paráfrase do lema "Entender o que você busca para encontrar o que você precisa", da empresa de consultoria Price waterhouse.

mais um recurso importante para a gestão do conhecimento da organização. Porém, pode-se afirmar que a ausência de uma arquitetura semântica pode comprometer uma melhor expressividade do conhecimento registrado pela corporação, principalmente se esta tiver como atividades mais relevantes a pesquisa e o desenvolvimento para inovação.

Os recursos tecnológicos da Web Semântica viabiliza a construção de uma arquitetura desta natureza. Esta tecnologia converge para a implementação, os esforços de áreas de conhecimento como inteligência artificial, modelagem conceitual, representação, organização e representação do conhecimento. O desafio futuro será aplicar esses conhecimentos e tecnologia como efetivo apoio à gestão do conhecimento.

Com a arquitetura em questão dois tipos de escalabilidade estão presentes: uma interna à arquitetura e outra externa. A escalabilidade interna diz respeito à capacidade da arquitetura de agregar novas estruturas e novos conteúdos sem perder sua identidade holística. É como a técnica de plantio direto<sup>44</sup>, onde o novo se integra ao que é velho em perfeita harmonia. A segunda escalabilidade é a possibilidade de composição de conhecimentos por meio da agregação com outros conhecimento de igual estrutura de representação. Ou seja, não é preciso reinventar conceitos sob mesmos contextos e condições de significância. A reinvenção deve ocorrer quando diz respeito a novas perspectivas ou novidade de conhecimento. Assim, se o conceito “Mandioca” está bem definido na DBPedia (Anexo I) basta apenas utilizá-lo agregando-o ao novo conhecimento que se está formulando. Não há necessidade nem esforço de recriar o conceito.

Além disso, as possibilidades de compartilhamento se ampliam de forma quantitativa e qualitativa. De forma quantitativa porque, considerando o ambiente sob um único padrão tecnológico de estruturas e dados, a rede semântica interna da corporação poderá se integrar e interoperar com as redes semânticas externas de outras corporações, instituições ou entidades sociais de qualquer natureza. Por outro lado, o compartilhamento do conhecimento se amplia qualitativamente porque conhecimento intercambiado pode ser melhor validado pelas instituições

---

44 [http://pt.wikipedia.org/wiki/Plantio\\_direto](http://pt.wikipedia.org/wiki/Plantio_direto)

competentes. O conceito “Mandioca” pode estar bem descrito na DBPedia, a qual compete uma boa descrição. Mas o mesmo conceito pode ser melhor caracterizado pela Unidade de Pesquisa Embrapa Acre quando se referir ao contexto local dessa região da Amazônia com suas idiossincrasias particulares. Da mesma forma, uma outra Unidade de Pesquisa como a Embrapa Mandioca Fruticultura poderá apresentar, em sua competência, o mesmo conceito sob seu olhar próprio, a partir de suas pesquisas e experiências. Ou, por sua vez, a Embrapa Sede, poderia dar uma visão mais holística, mais abrangente, em termos de território nacional, sobre a cultura da mandioca. Com esses dados semânticos ligados disponíveis um usuário ou aplicação poderia agregá-los de forma a criar sua própria perspectiva ou faceta.

Neste sentido, a árvore do conhecimento poderá ser constituída partir de várias fontes de conhecimento, cada uma delas operando segundo sua competência, assegurando que o conhecimento seja criado, mantido e utilizado por diversas entidades, cada uma com sua experiência, vivência e dinâmica, sob uma ampla gestão do conhecimento, que é tecido em conjunto. E isto é operar sobre a complexidade.

## 4.2 Perspectivas para trabalhos futuros

O campo da aplicação da Web Semântica em gestão do conhecimento ainda está em desenvolvimento. É preciso avaliar melhor o papel da tecnologia nas estruturas e processos informacionais da corporação, bem como os impactos sobre os usuários e outros agentes envolvidos. Processos de desenvolvimento de arquitetura de dados semânticos precisam ser desenvolvidos, aperfeiçoados e validados.

A integração de fontes heterogêneas de dados semânticos ligados poderá ser uma evolução futura para os atuais *data warehouses* baseados nos modelos relacionais de tabelas e registros ligados, onde a representação do conhecimento sob a rígida estrutura bidimensional dos dados pode evoluir para a estrutura fractal e em rede dos dados ligados.



Em termos de conhecimento explícito boa parte dos processos de gestão do conhecimento são voltados para o compartilhamento de conteúdos em documentos, para os mecanismos de comunicação que promovam a interatividade, e para a mineração de dados em bases de dados relacionais consolidadas a partir de fontes heterogêneas de dados estruturados, semi-estruturados ou não-estruturados. É preciso avaliar melhor os processos de gestão do conhecimento a partir do uso de dados semânticos ligados.

No âmbito da pesquisa agropecuária é possível desenvolver, testar e validar processos de integração envolvendo dados semânticos em conjunto com as fontes de informação tradicionais como bancos de dados, arquivos de dados, documentos eletrônicos anotados e conteúdos institucionais na *Web*.

## REFERÊNCIAS

ABBAGNANO, Nicola. Dicionário de Filosofia. São Paulo: Martins Fontes, 2007.

ABECKER, Andreas; BERNARDI, Ansgar; HINKELMANN, Knut; KUHN, Otto; SINTEK, Michael. Toward a technology for organizational memories. **Journal IEEE Intelligent Systems**, v. 13, n. 3, p. 40-48, 1998. Disponível em: <<http://www.dfki.uni-kl.de/~sintek/Papers/IEEE/OM-Technology.pdf.gz>>. Acesso em: 12 jun. 2012.

ALLEMANG, Dean. Semantic web and the linked data enterprise. In: WOOD, David. (Ed.). **Linking Enterprise Data**. New York: Springer-Verlag, 2010.

ALVARES, Lillian. **Organização da Informação**. Notas de aula da disciplina "Fundamentos em Organização da Informação". Disponível em: <<http://www.alvarestech.com/lillian/Fundamentos/Modulo1/Aula12OI.pdf>>. Acesso em: 26 jan. 2013.

AN, Yuan; BORGIDA, Alex; MYLOPOULOS, John. Discovering the semantics of relational tables through mappings. **Journal on Data Semantics**, n. 7, LNCS 4244, p. 1-32. 2006. Disponível em: <[http://www.cs.toronto.edu/semanticweb/maponto/papers/JoDS\\_VII.discovering.pdf](http://www.cs.toronto.edu/semanticweb/maponto/papers/JoDS_VII.discovering.pdf)> Acesso em: 26 jun. 2012.

ANDERSON, Philip W. More is different: broken symmetry and the nature of the hierarchical structure of science. **Science**, v. 177, n. 4047, p. 393-396, 1972.

ANTONIOU, Grigoris; HARMELEN, Frank Van. **A semantic web primer**. 2. ed. Massachusetts: Massachusetts Institute of Technology, 2008. 264 p.

BAADER, Franz; GANTER, Bernhard; SATTLER, Ulrike; SERTKAYA, Baris. Completing Description Logic Knowledge Bases using Formal Concept Analysis. In: IJCAI'07 PROCEEDINGS OF THE 20TH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, San Francisco, CA, USA: Morgan Kaufmann, 2007. p. 230-235.

BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (Org.). **Educação, universidade e pesquisa**. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. p.35-60.

BATESON, Gregory. **Steps to an ecology of mind: collected essays in anthropology, psychiatry, evolution, and epistemology**. New Jersey: Jason Aronson, 1987. 361 p.

BELKIN, N. J., ROBERTSON, S. E. Information science and the phenomenon of information. **Journal of the American Society for Information Science**, v.27, n.4, p. 197-204, Jul-Aug., 1976.

BERNERS-LEE. **Information Management: a proposal**. Genebra, mar. 1989. Disponível em: <<http://www.w3.org/History/1989/proposal.html>>. Acesso em: 9 fev. 2013.

\_\_\_\_\_. **Linked data**. 2006. Disponível em <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 7 fev. 2013.

\_\_\_\_\_. **Semantic web concepts**. Bio-IT World, 2005. Disponível em: <<http://www.w3.org/2005/Talks/0517-boit-tbl/>>. Acesso em: 4 fev. 2013.

\_\_\_\_\_. ; HENDLER, James; LASSILA, ORA. The semantic web. **Scientific American**, p. 29-37. May 2001.

BLOEHDORN, Stephan; HASSE, Peter; HUANG, Zhisheng; SURE, York; VOLKER, Johanna; HARMELEN, Frank va; STUDER, Rudi. Ontology Management. In: DAVIES, John; GROBELNIK, Marko; MLADENIC, Dunja. (Ed.). **Semantic Knowledge management: integrating ontology management. Knowledge discovery and human language technologies**. Berlin: Springer-Verlag, 2009. p. 3-20.

BRABANDERE, Luc de. **O lado oculto das mudanças**. Rio de Janeiro: Campus, 2006.

BREITMAN, K. **Web semântica: a internet do futuro**. Rio de Janeiro: LTC, 2005.

BOHR, Niels Henrik David. **Física atômica e conhecimento humano - Ensaios 1932-1957**. 4. reimpr., Rio de Janeiro: Contraponto, 2008. 129 p. Traduzido do original: Atomic physics and human knowledge. Copenhagen: Niels Bohr Archive, 1958.

BUCKLAND, Michael. Information as thing. In: **Journal of the American Society of Information Science**, v. 42, n. 5, p. 351–360, 1991. Disponível em: <<http://www.publicpraxis.com/wp-content/uploads/2011/01/informationasthing.pdf>>. Acesso em: 5 jun. 2012.

\_\_\_\_\_, Michael. What is a document? **The Journal of the American Society of Information Science**, v. 48, n. 9, p. 804-809, 1997. Disponível em: <[http://polaris.gseis.ucla.edu/gleazer/260\\_readings/Buckland.pdf](http://polaris.gseis.ucla.edu/gleazer/260_readings/Buckland.pdf)>. Acesso em: 5 jun. 2012.

BUSH, Vannevar. As we may think. **Atlantic Magazine**, 1945. Disponível em: <<http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881/>>. Acesso em: 5 jun. 2012.

CAPRA, Fritjof. **O ponto de mutação: a ciência, a sociedade e a cultura emergente**. 25. ed. São Paulo: Cultrix, 1982.

CAPURRO, Rafael; HJORLAND, Birger. O conceito de informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.12, n.1, p. 148-207. jan./abr., 2007. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-99362007000100012](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362007000100012)>. Acesso em: 11 jun. 2012.

CASANAVE, Cory. **Designing a semantic repository. Integrating architectures for reuse and integration**. May, 2007. Disponível em: <<http://www.w3.org/2007/06/eGov-dc/papers/SemanticRepository.pdf>>. Acesso em: 10 Jan. 2013.

CERVO, Amado Luiz Cervo; BERVIAN, Pedro Alcino. **Metodologia científica**. São Paulo: Makron Books, 1996.

CHOO, Chun Wei. Information management for the intelligent organization. 3. ed. **American Society for Information Science and Technology**. Medford: Information Today, 2002. 325 p.

\_\_\_\_\_. **The knowing organization**: how organizations use information to construct meaning, create knowledge, and make decisions. 2. ed. New York: Oxford, 2006. 354 p.

CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de Oliveira. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos, 2008. 451 p.

DAHLBERG, Ingetraut. **Knowledge organization**, 2006. Disponível em: <[http://www.iva.dk/bh/lifeboat\\_ko/CONCEPTS/knowledge\\_organization\\_Dahlberg.htm](http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_Dahlberg.htm)>. Acesso em: 27 jan. 2013.

\_\_\_\_\_. **Teoria do conceito**. Tradução para o português do Prof. Astério Tavares Campos. Ciência da Informação, Brasília, v. 7, n. 2, p. 101-107, 1978. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/viewFile/1680/1286>>. Acesso em: 28 jan. 2013.

DEMO, Pedro. **Metodologia do Conhecimento Científico**. São Paulo: Atlas, 2000. 216 p.

\_\_\_\_\_, Pedro. **Praticar ciência**: metodologias do conhecimento científico. São Paulo: Saraiva, 2011. 208 p.

DIENG, Rose; CORBY, Olivier; GIBOIN, Alain; RIBIÈRE, Myriam. **Methods and tools for corporate knowledge management**. Institut National de Recherche en Informatique et en Automatique, Septembre, 1998. Disponível em: <<http://hal.inria.fr/docs/00/07/32/03/PDF/RR-3485.pdf>>. Acesso em: 12 fev. 2013.

\_\_\_\_\_. Corporate semantic webs. In: SCHWARTZ, David G. (Ed.). **Encyclopedia of Knowledge Management**. London: Idea Group Reference, 2005. p. 67-80.

\_\_\_\_\_. MATTA, Nada. **Knowledge management and organizational memories**. Norwell, MA, USA: Kluwer Academic Publishers, 2002.

DOLBEAR, Catherine. **W3C semantic web use cases and case studies. Case study: semantic web technology at ordnance survey.** March, 2007. Disponível em: <<http://www.w3.org/2001/sw/sweo/public/UseCases/OrdSurvey/>>. Acesso em: 11 Jan. 2013.

DOSI, Giovanni. The nature of the innovative process. In: DOSI, Giovanni; FREEMAN, C.; NELSON, R.; SILVERBERG, G.; SOETE, L. (Ed.). **Technical change and economic theory**, London: Pinter Publishers, London. 1988. cap. 10, p. 221-238.

DRETSKE, Fred I. **Knowledge and the flow of information.** Stanford, California: CSLI Publications. 1999. The David Hume Séries of Philosophy and Cognitive Science Reissues.

ELST, Ludger van; ABECKER, Andreas. Ontologies for information management: balancing formality, stability, and sharing scope. **Expert Systems with Applications**, n. 23, p. 357-366, 2002.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **American Association for Artificial Intelligence, AI Magazine**, v. 17, n. 3, 1996. Disponível em: <<http://www.aaai.org/ojs/index.php/aimagazine/article/download/1230/1131>>. Acesso em: 22 jun. 2012.

FENSEL, Dieter. **Ontologies: A silver bullet for knowledge management and electronic commerce.** 2. ed., New York: Springer-Verlag, 2004.

FLORIDI, Luciano. Semantic conceptions of information. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**, 2011. Disponível em: <<http://plato.stanford.edu/entries/information-semantic/>>. Acesso em: 29 jun. 2012.

FRAWLEY, William J.; PIATETSKY-SHAPIRO, Gregory; MATHEUS, Christopher J. **Knowledge Discovery in Databases: An Overview.** AI Magazine, v. 13, n. 3, 1992. Disponível em: <<http://www.aaai.org/ojs/index.php/aimagazine/article/download/1011/929>>. Acesso em: 22 jun. 2012.

FROMM, Jochen. **The emergence of complexity.** Kassel, Germany: Kassel

University Press, 2004, 208 p.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 1999. 207 p.

GRAY, Jim. Scientific data management in the coming decade. **ACM SIGMOD Record**, v. 34, n. 4, p. 34-41, 2005.

GOLDSTEIN, Jeffrey. Emergence as a construct: history and issues. **Emergence-Journal of Complexity Issues in Organizations and Management**, v. 1, n. 1, p. 49-72, 1999.

GÓMEZ-PÉREZ, Asunción; FERNÁNDEZ-LÓPEZ, Mariano.; CORCHO, Oscar. **Ontological engineering**: with examples from the areas of knowledge management, ecommerce and the semantic web. London: Springer, 2004. 411p.

GRUBER, Thomas R. **A translation approach to portable ontology specifications**. Stanford University, 1993. Disponível em : <<http://tomgruber.org/writing/ontologia-kaj-1993.pdf>>. Acesso em: 13 jan. 2013.

\_\_\_\_\_. Toward principles for the design of ontologies used for knowledge sharing. **International Journal of Human and Computer Studies**, v. 43, n. 5/6, p. 907-928, 1995.

\_\_\_\_\_. Ontology. In: LIU, Ling; OZSU, M. Tamer. (Ed.). **The Encyclopedia of Database Systems**. New York: Springer-Verlag, 2009. Disponível em: <<http://tomgruber.org/writing/ontology-definition-2007.htm>>. Acesso em: 13 Jan. 2013.

GLUCK, Mike. Exploring the relationship between user satisfaction and relevance in information systems. In: **Information Processing & Management**, v. 32, n. 1, p. 89-104, 1996.

GUARINO, Nicola. Formal ontology and information systems. In: \_\_\_\_\_. **Formal ontology in information systems**. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam: IOS Press, p. 3-15, 1998.

\_\_\_\_\_. Semantic Matching: Formal ontological distinction for information organization, extraction, and integration. In: PAZIENZA, Maria Teresa. (Ed.). **Information extraction: A multidisciplinary approach to na emerging information technology**. London: Springer-Verlag, 1997.

\_\_\_\_\_.; OBERLE, Daniel; STAAB. What is an ontology. In: STAAB, S.; STUDER, R. (Ed.). **Handbook on ontologies**. 2. ed.. Berlim: Springer-Verlag, 2009. p. 1-17.

HEATH, Tom; BIZER, Christian. **Linked Data: evolving the web into a global data space**. Palo Alto, California: Morgan & Claypool Publishers, 2011. 121 p.  
HEBELER, John; FISHER, Matthew; BLACE, Ryan; PEREZ-LOPEZ, Andrew. **Semantic web programming**. Indianapolis: Wiley, 2009.

HINTON, Andrew. The machineries of context: new architectures for a new dimension. **Journal of Information Architecture**. v. 1. n. 1. p. 37-47. 5 May 2009. Disponível em: <<http://journalofia.org/volume1/issue1/04-hinton/jofia-0101-04-hinton.pdf>>. Acesso em: 7 jun. 2012.

HJORLAND, Birger. Controlled vocabulary (or prescriptive vocabulary). In: \_\_\_\_\_. **Lifeboat for Knowledge Organization**, 2007. Disponível em: <[http://www.iva.dk/bh/Lifeboat\\_KO/CONCEPTS/controlled\\_vocabulary.htm](http://www.iva.dk/bh/Lifeboat_KO/CONCEPTS/controlled_vocabulary.htm)>. Acesso em 27 jan. 2013.

\_\_\_\_\_. Semantic relations (meaning relations). In: \_\_\_\_\_. **Lifeboat for Knowledge Organization**, 2007. Disponível em: <[http://www.iva.dk/bh/lifeboat\\_ko/CONCEPTS/semantic\\_relations.htm](http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/semantic_relations.htm)>. Acesso em 27 jan. 2013.

\_\_\_\_\_. Knowledge organization systems (KOS). In: \_\_\_\_\_. **Lifeboat for Knowledge Organization**, 2008. Disponível em: <[http://www.iva.dk/bh/lifeboat\\_ko/CONCEPTS/knowledge\\_organization\\_systems.htm](http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm)>. Acesso em 27 jan. 2013.

\_\_\_\_\_, Birger. Semantics and knowledge organization. **Annual Review of Information Science and Technology**, v. 41, n. 1, 2008.



\_\_\_\_\_, Birger. What is Knowledge Organization (KO)? **Knowledge organization. International Journal devoted to Concept Theory, Classification, Indexing and Knowledge Representation**, v. 35, n. 2/3, 2008. Disponível em: <[http://arizona.openrepository.com/arizona/bitstream/10150/106183/1/What\\_is\\_Knowledge\\_Organization\\_Finalrev\\_correctedC.doc](http://arizona.openrepository.com/arizona/bitstream/10150/106183/1/What_is_Knowledge_Organization_Finalrev_correctedC.doc)>. Acesso em: 28 jun. 2012.

HODGE, Gail. **Systems of knowledge organization for digital libraries: Beyond Traditional Authority Files**. 2000. Washington, DC: The Council on Library and Information Resources. Disponível em <<http://old.diglib.org/pubs/dlf090/dlf090.pdf>>. Acesso em 27 jan. 2013.

HOUAISS, Antônio; VILLAR, Mauro de Salles; FRANCO, Francisco Manoel de Mello. **Dicionário da língua portuguesa**. Rio de Janeiro: Objetiva, 2009.

HYDE, Kenneth F. Recognising deductive processes in qualitative research. **Qualitative Market Research: An International Journal**, v. 3, n. 2, p. 82-89. 2000.

INMON, William H. **Building the data warehouse**. 3. ed. New York: Wiley, 2002. 428 p.

\_\_\_\_\_; NESAVICH, Anthony. **Tapping into unstructured data - integrating unstructured data and textual analytics into business intelligence**. New York: Prentice Hall, 2008.

\_\_\_\_\_; O'NEIL, Boonie; FRYMAN, Lowell. **Business metadata: capturing enterprise knowledge**. New York: Morgan Kaufmann. 2008.

JOHNSON, Steven. **O mapa fantasma: como a luta de dois homens contra o cólera mudou o destino de nossas metrópoles**. Rio de Janeiro: Jorge Zahar, 2008.

KAHANE, Adam. **Como resolver problemas complexos: uma forma aberta de falar, escutar e criar novas realidades**. São Paulo: Editora Senac, 2008. 208 p.

KASHYAP, Vipul; BUSSLER, Christoph; MORAN, Matthew. **The semantic web: Semantics for Data and Services on the Web**. Berlin: Springer, 2008. 414 p.

KIRYAKOV, Atanas; POPOV, Borislav; KITCHUKOV, Ilian; ANGELOV, Krasimir. Shared ontology for knowledge management. In: DAVIES, John; GROBELNIK, Marko; MLADENIC, Dunja. (Ed.). **Semantic knowledge management: integrating ontology management**. Knowledge discovery and human language technologies. Berlin: Springer-Verlag, 2009. p. 61-83.

KIRYAKOV, Atanas; DAMOVA, Mariana. Storing the semantic web: Repositories. In: DOMINGUE, John; FENSEL, Dieter; HENDLER, James A. (Ed.). **Handbook of semantic web technologies**. Berlin: Springer-Verlag, 2011. p. 233-297.

KOBASHI, Nair Yumiko. **Vocabulário controlado: estrutura e utilização**. Brasília: ENAP – Escola Nacional de Administração Pública, 2008. Disponível em: <[http://www2.enap.gov.br/rede\\_escolas/arquivos/vocabulario\\_controlado.pdf](http://www2.enap.gov.br/rede_escolas/arquivos/vocabulario_controlado.pdf)>. Acesso em: 27 jan. 2013.

LACY, Lee W. **OWL: representing information using the web ontology language**. Victoria BC, Canada: Trafford Publishing, 2005.

LANKHORST, Marc; et al. **Enterprise Architecture at Work: Modelling, Communication, and Analysis**. Berlin: Springer-Verlag, 2005.

LASSILA, Ora; MCGUINNESS, Deborah. The role of frame-based representation on the semantic web. In: **ETAI (Electronic Transactions in Artificial Intelligence)**, v. 6, 2001. Disponível em: <<http://www.ksl.stanford.edu/people/dlm/etai/lassila-mcguinness-fbr-sw.html>>. Acesso em: 25 jan. 2013.

LE COADIC, Y. F. **A ciência da informação**. 2. ed. Brasília: Briquet de Lemos, 2004.

LENZERINI, Maurizio. Data integration: a theoretical perspective. In: PROCEEDING PODS '02. ACM SIGMOD-SIGACT-SIGART SYMPOSIUM ON PRINCIPLES OF DATABASE SYSTEMS. 21. Proceedings. 2002. p. 233-246.

LEVY, Alon Y. Logic-based techniques in data integration. In: MINKER, Jack (Ed.). **Logic-based artificial intelligence**. Norwell, MA, USA: Kluwer Publishers, November 2000. cap. 1, p. 1-27. Disponível em: <<http://www.csd.uoc.gr/~hy562/Papers/levy-di00.pdf>>. Acesso em: 26 jun. 2012.

LEVY, David M. Heroic measures: reflections on the possibility and purpose of digital preservation. In: ACM CONFERENCE ON DIGITAL LIBRARIES, 3. 1998. Proceedings. p. 152-161.

LEVY, Pierre. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. São Paulo: Editora 34, 1997.

LUDASCHER, Bertram; LIN, Kai; BOWERS, Shawn; JAEGER-FRANK, Efrat; MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de Metodologia Científica**. 5. ed. São Paulo: Atlas, 2003. 311 p.

MCINERNEY, Clarie Regina. Knowledge management and the dynamic nature of knowledge. **Journal of the American Society for Information Science and Technology**. Special Issue: Special Topic Issue: Knowledge Management. v. 53, n. 12, p. 1009-1018, 2002. Disponível em: <[http://comminfo.rutgers.edu/~clairemc/KM\\_dynamic\\_nature.pdf](http://comminfo.rutgers.edu/~clairemc/KM_dynamic_nature.pdf)>. Acesso em: 5 jun. 2012.

MIZZARO, Stefano. Relevance: The whole history. **Journal of The American Society for Information Science**, v. 48, n. 9, p. 810-832, 1997.

MOREIRA, Marco Antonio. Mapas conceituais e aprendizagem significativa. **Cadernos de Aplicação**, Porto Alegre, v. 11, n. 2, p. 143-156, 1998. Disponível em: <<http://www.if.ufrgs.br/~moreira/mapasport.pdf>>. Acesso em: 14 jan. 2013.

MORIN, Edgar. O método 1 - a natureza da natureza. 2. ed. Porto Alegre: Sulina, 2008. 479 p.

\_\_\_\_\_; LE MOIGNE, Jean-Louis. **Inteligência da complexidade: epistemologia e pragmática**. Lisboa: Instituto Piaget, 2009. 527 p.

\_\_\_\_\_. **Introdução ao pensamento complexo**. 4. ed. Porto Alegre: Sulina, 2011. 120 p.

MOSER, Paul K.; MULDER, Dwayne H.; TROUT, J. D. **A teoria do conhecimento: uma introdução temática**. 2. ed. São Paulo: Martins Fontes, 2009. 233 p.

MUELLER, Suzana Pinheiro Machado. A ciência, o sistema de comunicação científica e a literatura científica. In: CAMPELLO, B. S.; CENDÓN, B. V.; KREMER, J. M. (Org.). **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: UFMG, 2000.

\_\_\_\_\_. Suzana Pinheiro Machado. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v. 35, p. 925, 2006. Disponível em: <[www.scielo.br/pdf/ci/v35n2/a04v35n2.pdf](http://www.scielo.br/pdf/ci/v35n2/a04v35n2.pdf)>. Acesso em: 16 dez. 2011.

NEBOT, Victoria; BERLANGA, Rafael. Building data warehouses with semantic data. **Decision Support Systems**, v. 52, n. 4, p. 853-868, 2012.

NONAKA, Ikujiro. A dynamic theory of organization knowledge creation.

**Organization Science**, v. 5, n. 1, p. 14-37, 1994. Disponível em:

<<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.2590&rep=rep1&type=pdf>> Acesso em 25 jun. 2012.

\_\_\_\_\_.; TAKEUCHI, Hirotaka. **The knowledge-creating company**: how japanese

companies create the dynamics of innovation. New York: Oxford, 1995. 284 p.

\_\_\_\_\_.; TOYAMA, Ryoko; KONNO, Noboru. **SECI, ba and leadership**: a unified model of dynamic knowledge creation. *Long Range Planning*, v. 33, 2000. p. 5-34.

Disponível em: <<http://www.ai.wu.ac.at/~kaiser/literatur/nonaka-seci-ba-leadership.pdf>>. Acesso em: 24 jun. 2012.

NOVAK, Joseph D.; GOWIN, D. Bob. **Learning how to learn**. New York: Cambridge University Press, 1984. 199 p.

\_\_\_\_\_. **Learning, Creating, and Using Knowledge**: Concept maps as facilitative tools in schools and corporations. New York: Routledge, 1998, 317 p.

O'CONNOR, Timothy. Emergent properties. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**, 2012. Disponível em: <<http://plato.stanford.edu/entries/properties-emergent/>>. Acesso em: 22 jan. 2013.

OGBUJI, C.; BLACKSTONE; PIERCE, C. **W3C semantic web use cases and case studies. Case study**: a semantic web content repository for clinical research. October, 2007. Disponível em: <<http://www.w3.org/2001/sw/sweo/public/UseCases/ClevelandClinic/>>. Acesso em: 11 jan. 2013.

O'LEARY, Daniel.E. Enterprise knowledge management. **Journal Computer**, v. 31 n. 3, p. 54-61, mar. 1998.

POLANYI, Michael. **The tacit dimension**. First published Doubleday & Co, 1966. Gloucester, Mass: Reprinted Peter Smith, 1983.

POPPER, Karl. O incremento do conhecimento científico. 1960. In: MILLER, David. (Org.). **Textos escolhidos - Karl Popper**. Rio de Janeiro: Contraponto: Ed. PUC-Rio, 2010. cap. 3. p. 169-177.

\_\_\_\_\_, Karl. **A lógica da pesquisa científica**. Tradução de Leônidas Hegemberg. São Paulo: Editora Cultrix, 1975.

\_\_\_\_\_. **Conhecimento objetivo**. Belo Horizonte: Itatiaia, 1999. 394 p. Coleção Espírito do Nosso Tempo, v. 13. Traduzido da edição de 1973, corrigida. Objective Knowledge. London: Oxford University Press.

POWERS, Shelley. **Practical RDF**. Sebastopol, California: O'Reilly, 2003.

PRICE, Rosanne J.; SHANKS, Graeme. A Semiotic Information Quality Framework. In: DECISION SUPPORT IN AN UNCERTAIN AND COMPLEX WORLD: THE IFIP TC8/WG8.3 INTERNATIONAL CONFERENCE 2004. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.9817&rep=rep1&type=pdf>>. Acesso em: 18 jan. 2013.

QUIVY, Raymond; CAMPENHOUDT, Luc Van. **Manual de investigação em ciências sociais**. Lisboa: Gradiva, 1995.

ROSNER, Dietmar; GROTE, Brigitte; HARTMANN, Knut; HOFLING, Bjorn. From natural language documents to sharable product knowledge: a knowledge engineering approach. **Journal of Universal Computer Science**, v. 3, n. 8, 1997. Disponível em: <[http://www.jucs.org/jucs\\_3\\_8/from\\_natural\\_language\\_documents/Roesner\\_D.html](http://www.jucs.org/jucs_3_8/from_natural_language_documents/Roesner_D.html)>. Acesso em: 21 jun. 2012.

ROWLEY, Jennifer; HARTLEY, Richard. **Organizing knowledge: an introduction to managing access to information**. 4. ed. Aldershot, Hampshire, England: Ashgate, 2008.

SANTAELLA, Lúcia. **Comunicação e pesquisa: projetos para mestrado e doutorado**, São Paulo: Hacker Editores, 2001.

\_\_\_\_\_. **Semiótica aplicada**. São Paulo: Pioneira Thomson Learning, 2002. 186p.

SANTOS, Gil. Emergência: da mereologia à organização. In: DUQUE, David F. ; PAREJO, Emilio G.-C.; ANTÓN, Ignacio H. (Ed.). **Estudios de lógica, lenguaje y epistemología, IV Jornadas Ibéricas**, España: Fénix, 2010, p. 327-356. Disponível em: <[http://www.academia.edu/1105443/Emergencia\\_da\\_Mereologia\\_a\\_Organizacao](http://www.academia.edu/1105443/Emergencia_da_Mereologia_a_Organizacao)>. Acesso em 24 dez. 2012.

SARACEVIC, Tefko. The concept of "Relevance" in information science: a historical review. In: SARACEVIC, Tefko. (Ed.). **Introduction to information science**. New York: R. R. Bowker, 1970. p. 111-151.

SCHAMBER, Linda; EISENBERG, Michael B.; NILAN, Michael S. A re-examination of relevance: toward a dynamic, situational definition. **Information Processing & Management**, v. 26, n. 6, p. 755-776, 1990.

SCHARMER, Otto. **Teoria U: como liderar pela percepção e realização do Futuro emergente**. Rio de Janeiro: Elsevier, 2010. 432 p.

SICILIA, Miguel-Angel; LYTRAS, Miltiadis D. (Ed.). **Metadata and semantics**. New York: Springer, 2009. 552 p.

SOUZA, Rosali Fernandez de. A classificação como interface da Internet. **DataGramZero**, Rio de Janeiro, v. 1, n. 2, abr. 2000. Disponível em: <[http://www.dgz.org.br/abr00/Art\\_01.htm](http://www.dgz.org.br/abr00/Art_01.htm)>. Acesso em: 26 jan. 2013

\_\_\_\_\_. Organização do conhecimento. In: TOUTAIN, Lidia Maria Batista Brandão (Org.) **Para entender a Ciência da Informação**. Salvador: UDFBA, 2007. p. 101-121. ISBN 978-85-232-0477-8. (Coleção Sala de Aula 6) Disponível em: <<http://www.repositorio.ufba.br/ri/bitstream/ufba/145/1/Para%20entender%20a%20ciencia%20da%20informacao.pdf>>. Acesso em 26 jan. 2013.

SPEK, Rob van der Spek; SPIJKERVET, André. **Knowledge management: dealing intelligently with knowledge**. 4. ed. Netherlands: CIBIT Consultants, 2005. Disponível em: <[http://www.dnv.com/binaries/CIBIT\\_Knowledge\\_Management\\_Booklet\\_05\\_tcm4-353727.pdf](http://www.dnv.com/binaries/CIBIT_Knowledge_Management_Booklet_05_tcm4-353727.pdf)>. Acesso em: 12 fev. 2013.

SOWA, John F. **Conceptual structures: information processing in mind and machine**. The Systems Programming Series. Reading, Massachusetts: Addison-Wesley, 1984. 480 p.

\_\_\_\_\_. **Knowledge representation: logical, philosophical, and computational foundations**. Pacific Grove, California: Brooks/Cole, 2000a. 594p.

\_\_\_\_\_. Ontology, metadata, and semiotics. In: GANTER, Bernhard; MINEAU, Guy W. (Ed.). **CONCEPTUAL STRUCTURES: LOGICAL, LINGUISTIC, AND COMPUTATIONAL ISSUES. 8TH INTERNATIONAL CONFERENCE ON CONCEPTUAL STRUCTURES. PROCEEDINGS. Lecture Notes in Artificial Intelligence**, n. 1867, Berlin: Springer-Verlag, 2000b. p. 55-81.

STAAB, Steffen; SCHNURR, Hans-Peter; STUDER, Rudi; SURE, York. **Knowledge processes and ontologies**. In: **IEEE Intelligent Systems**. v. 16, n. 1, jan./feb. 2001. Special Issue on Knowledge Management.

SWANSON, Don R. On the fragmentation of knowledge, the connection explosion, and assembling other people's ideas. **Bulletin of the American Society for Information Science and Technology**, v. 27, n. 3, p. 12-14, feb./mar. 2001.

TAKEUCHI, Hirotaka; NONAKA, Ikujiro. **Gestão do conhecimento**. Porto Alegre:

Artmed, 2008. 319 p.

TAYLOR, Arlene; JOUDREY, Daniel N. **The organization of information**. 3. ed. Westport, Connecticut, USA: Libraries Unlimited, 2009.

TUOMI, Ilkka. Abstraction and history - from institutional amnesia to organizational memory. HAWAII ANNUAL INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, v. 4, p. 303-312. 1995.

\_\_\_\_\_. Data is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organizational memory. **Journal of Management Information Systems**, v. 16, n. 3, p 107-121, Fall 1999.

USCHOLD, Michael. **Ontologies ontologies everywhere – but who knows what to think?** Palestra apresentada no Protégé Users Conference, Stanford, CA, July 2006. Disponível em: <[http://protege.stanford.edu/conference/2006/submissions/slides/1.2\\_Uschold.pdf](http://protege.stanford.edu/conference/2006/submissions/slides/1.2_Uschold.pdf)>. Acesso em: 1 fev. 2013.

VICKERY, Brian C.; VICKERY, Alina. **Information science in theory and practice**. 3. ed.. München: K. G. Saur, 2004. 400 p.

W3C. Semantic Web Activity. SEMANTIC WEB KICK-OFF SEMINAR IN FINLAND, Nov 2, 2001. Disponível em: <<http://www.w3.org/2001/12/semweb-fin/w3csw>>. Acesso em: 6 fev. 2013.

\_\_\_\_\_. **RDF Vocabulary Description Language 1.0: RDF Schema**. W3C Recommendation 10, February 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>>. Acesso em: 2 jul. 2012.

\_\_\_\_\_. **Resource Description Framework (RDF): concepts and abstract syntax**. W3C Recommendation. 10 February 2004. Disponível em: <<http://www.w3.org/TR/rdf-concepts/>>. Acesso em: 4 jul. 2012.

\_\_\_\_\_. **SKOS Core Guide**. Editor's Draft 15 February 2005. Disponível em: <<http://www.w3.org/2004/02/skos/core/guide/2005-02-15.html>>. Acesso em: 7 fev. 2013.



\_\_\_\_\_. **Tutorial on Semantic Web Technologies**. Trento, Italy, on the 14th December, 2005. Disponível em: <<http://www.w3.org/2005/Talks/1214-Trento-IH/Overview.pdf>>. Acesso em: 03 fev. 2013.

\_\_\_\_\_. **Semantic Layer Cake**. 2007. Disponível em: <<http://www.w3.org/2007/03/layerCake.svg>>. Acesso em: 03 fev. 2013.

WHEELDON, Johannes P; AHLBERG, Mauri K. **Visualizing social science research: maps, methods, & meaning**. Thousand Oaks, California, USA: SAGE Publications, 2012. cap. 5, p. 113-148. Disponível em: <[http://www.uk.sagepub.com/upm-data/41670\\_5.pdf](http://www.uk.sagepub.com/upm-data/41670_5.pdf)>. Acesso em: 17 jan. 2013

WIEDERHOLD, Gio. Interoperation, mediation, and ontologies. INTERNATIONAL SYMPOSIUM ON FIFTH GENERATION COMPUTER SYSTEMS (FGCS94)ICOT, Tokyo, Japan, 1994.

WIIG, Karl M. Knowledge management: an Introduction and perspective. **Journal of Knowledge Management**, v. 1, n. 1, p. 6-14, 1997.

WOOD, David. **Linking enterprise data**. New York: Springer, 2010.

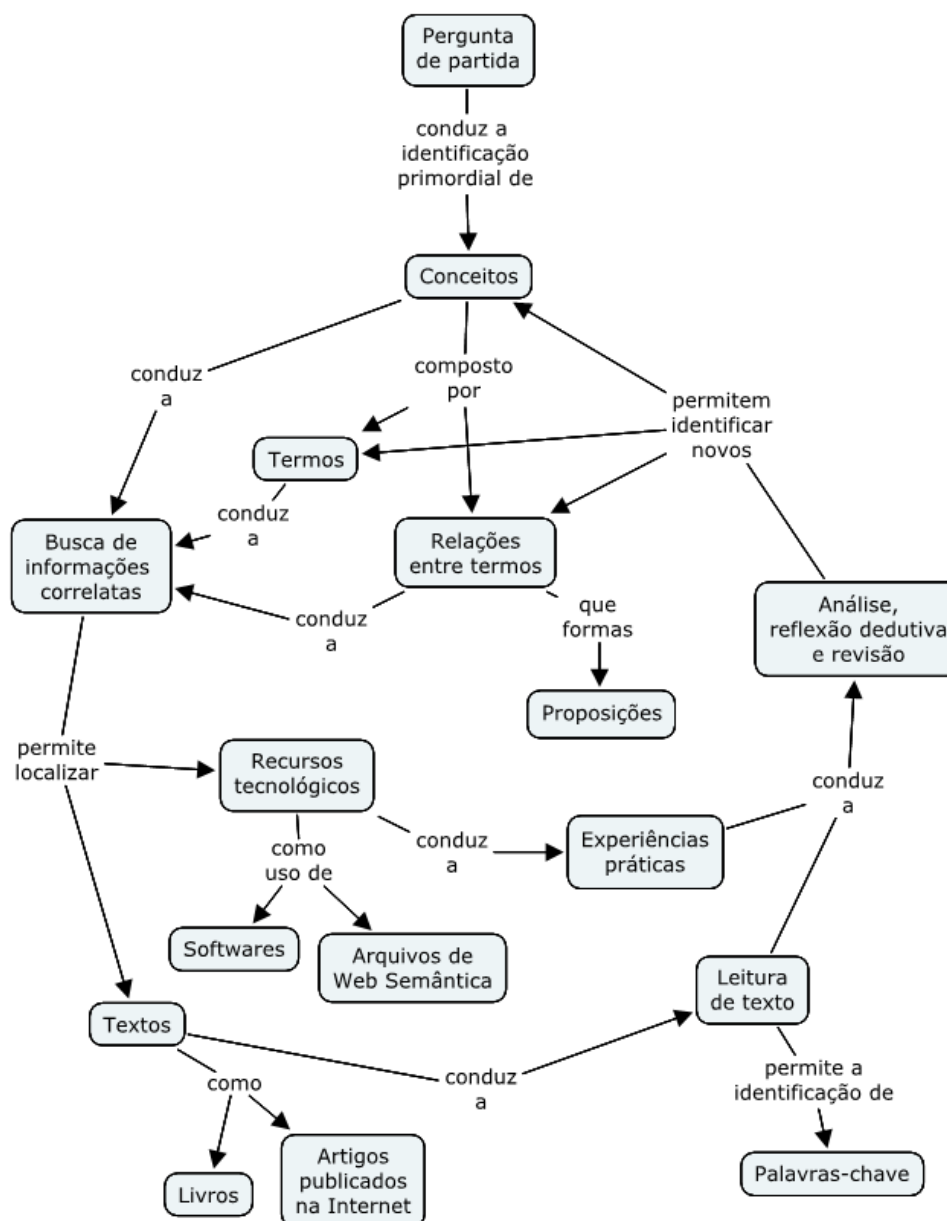
YU, Liyang, **A developer's guide to the semantic web**. Berlin: Springer, 2011. 608 p.

ZENG, Marcia Lei. Principles of controlled vocabularies. CONSTRUCTION OF CONTROLLED VOCABULARIES A PRIMER. Disponível em: <<http://marciazeng.slis.kent.edu/Z3919/2principle.htm>>. Acesso em: 31 jan. 2013.

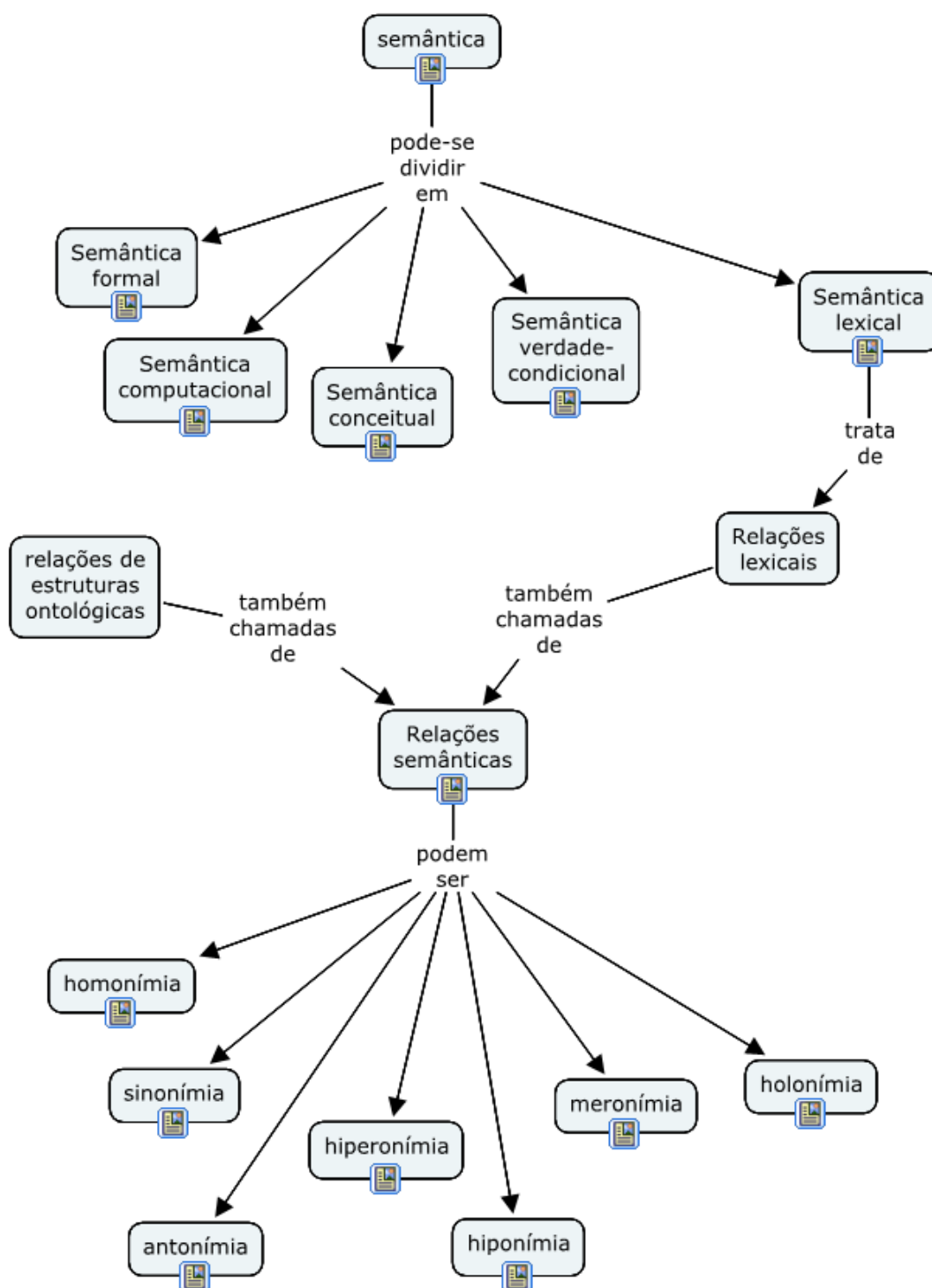
## APÊNDICES

### Apêndice A - Mapa conceitual do processo metodológico desta dissertação

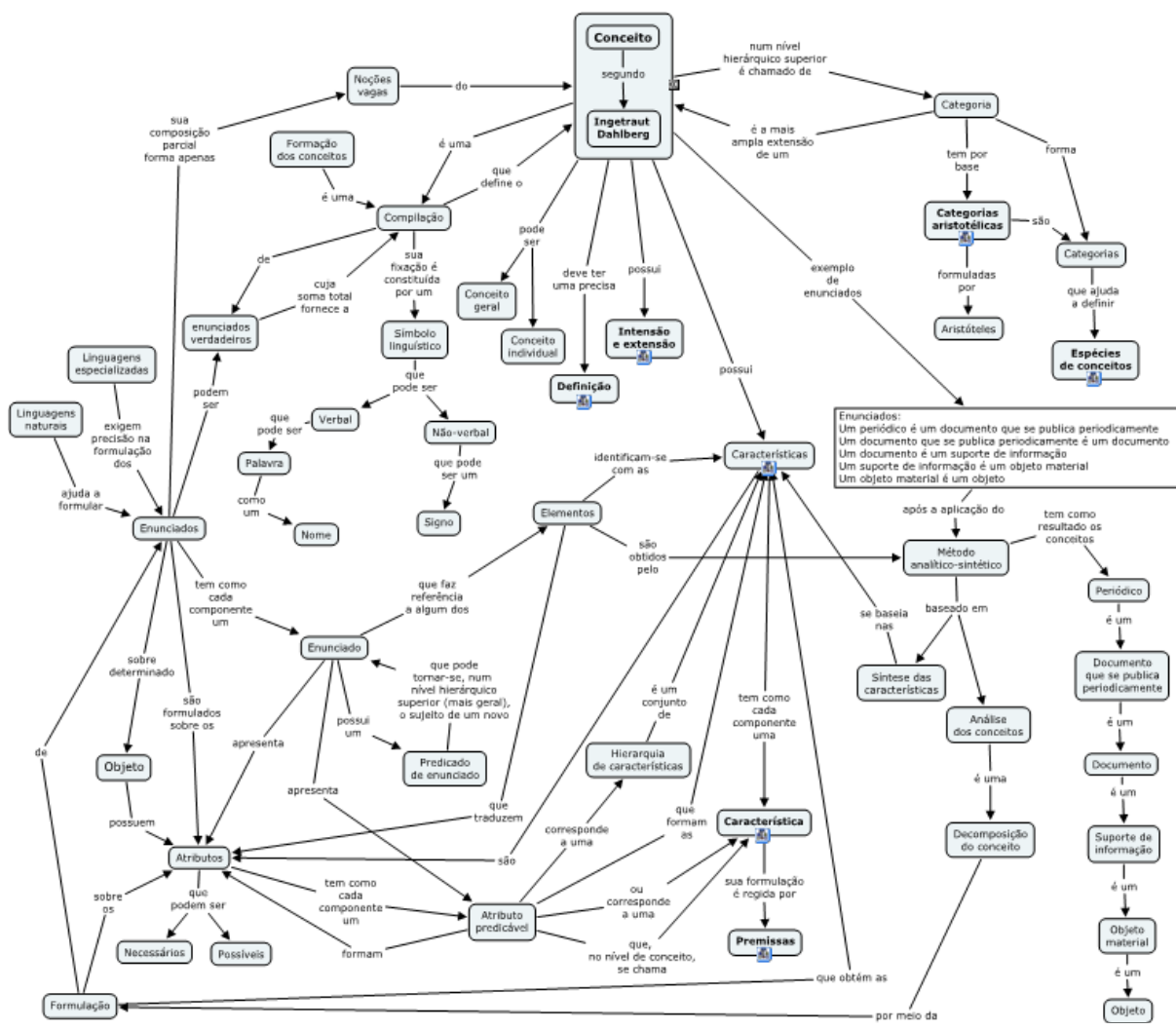
O mapa conceitual abaixo descreve a estratégia metodológica de busca de informações correlatas e deduções com base numa pergunta de partida. Trata-se de uma adaptação da metodologia de Quivy e Campenhoudt (1995, p. 27) para pesquisa exploratória descritiva pelo método dedutivo.



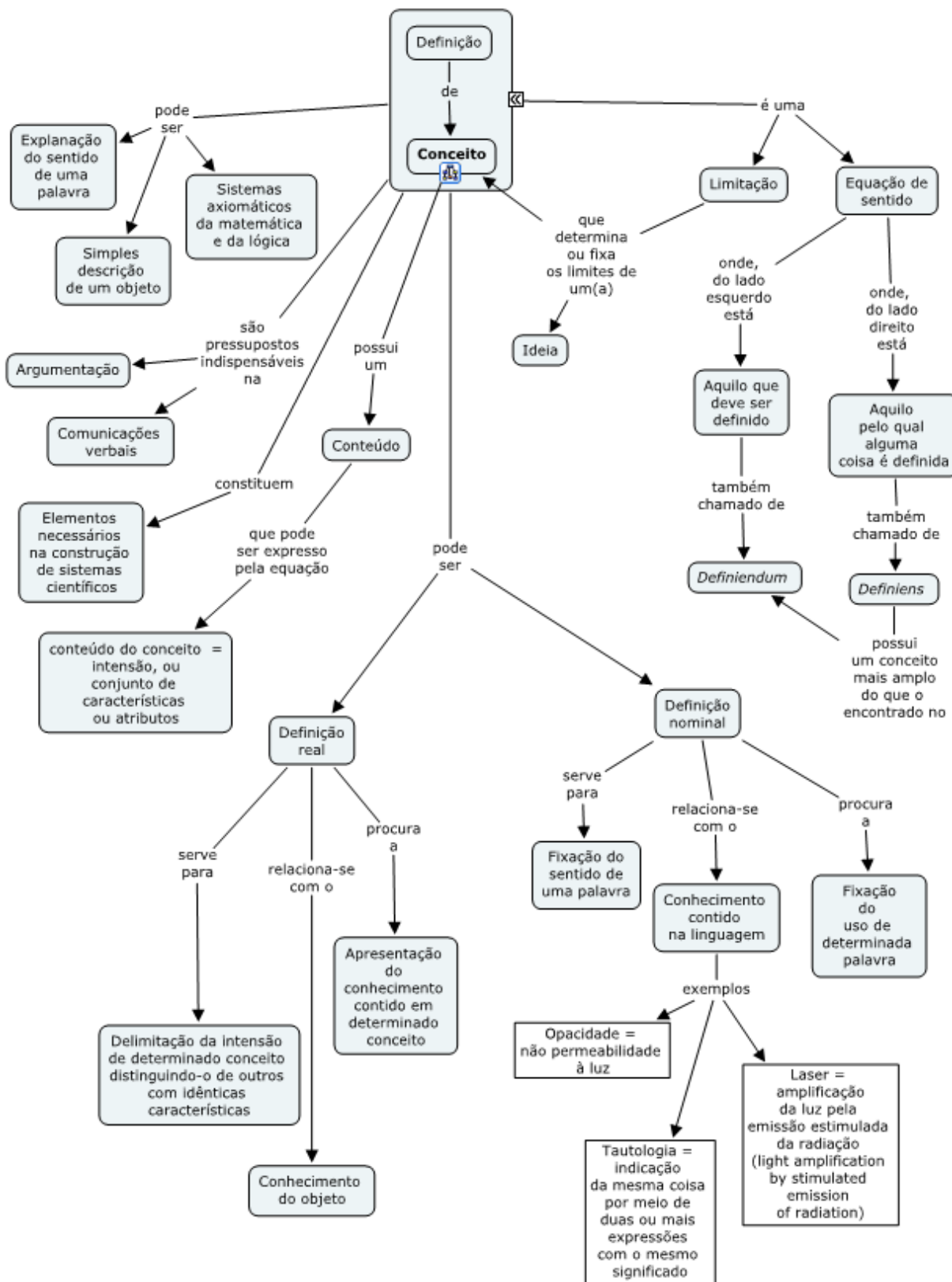
## Apêndice B - Mapa conceitual das relações semânticas



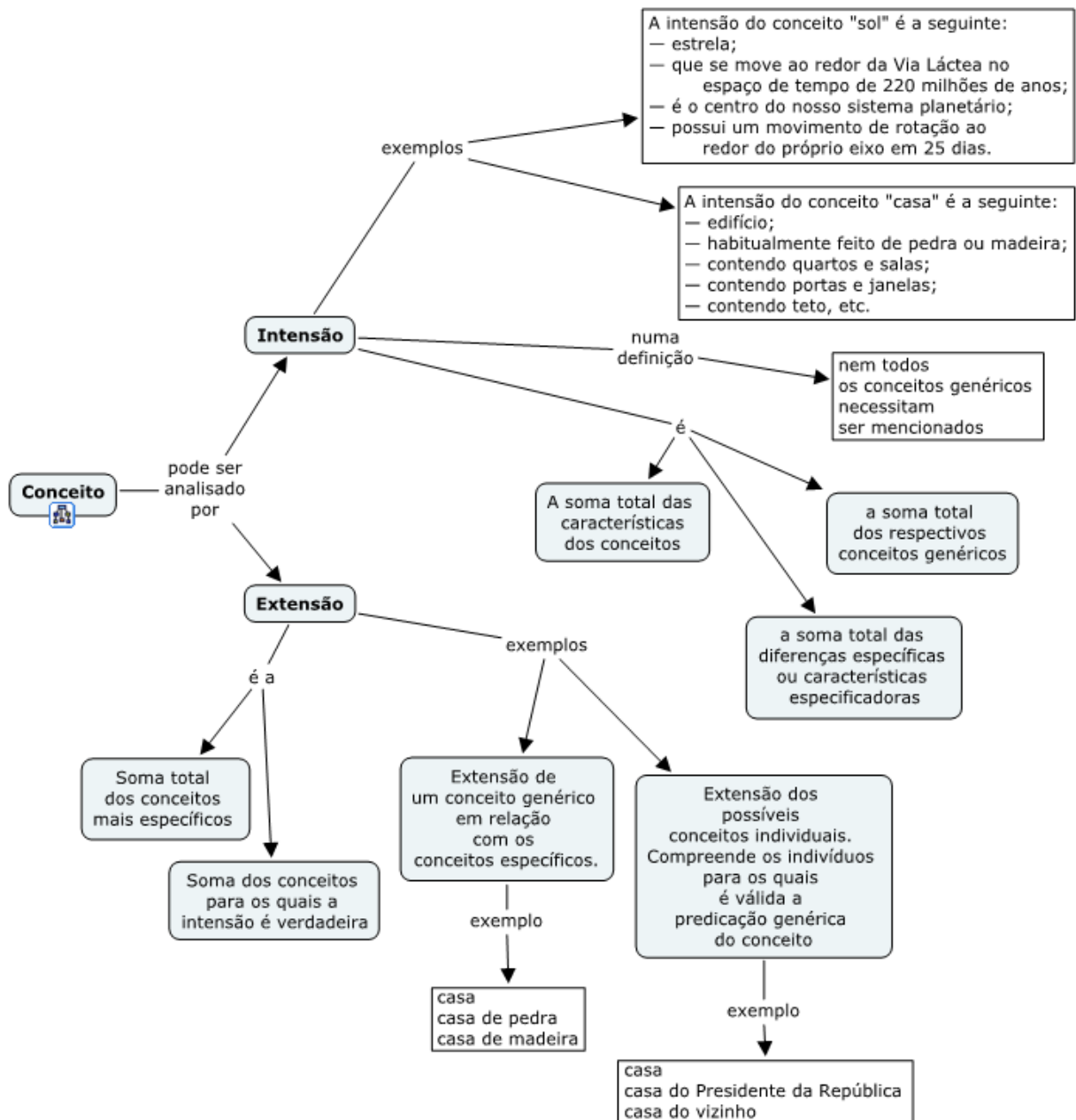
### Apêndice C - Mapa conceitual do processo de conceitualização



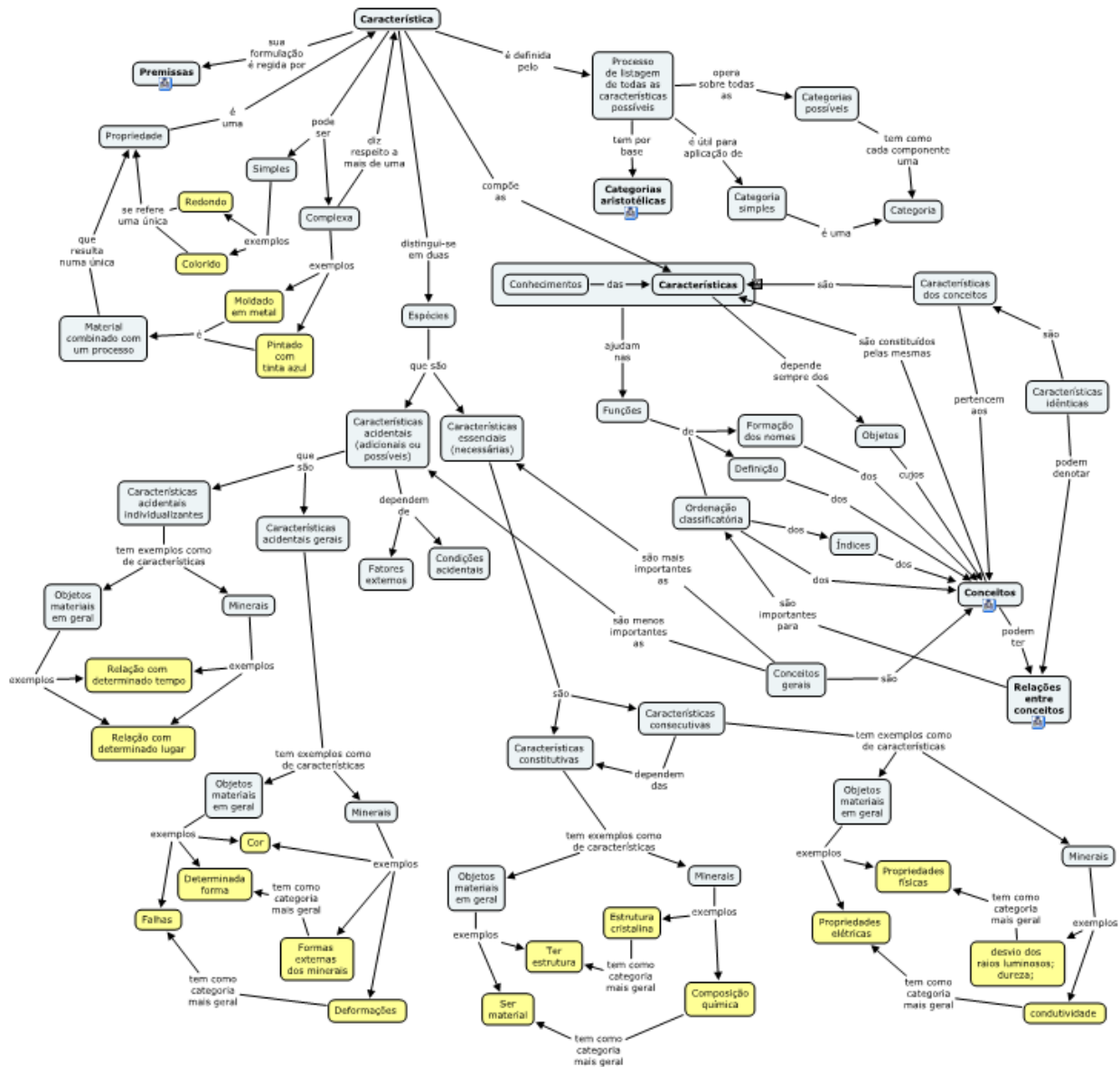
Apêndice D - Mapa conceitual da definição de um conceito



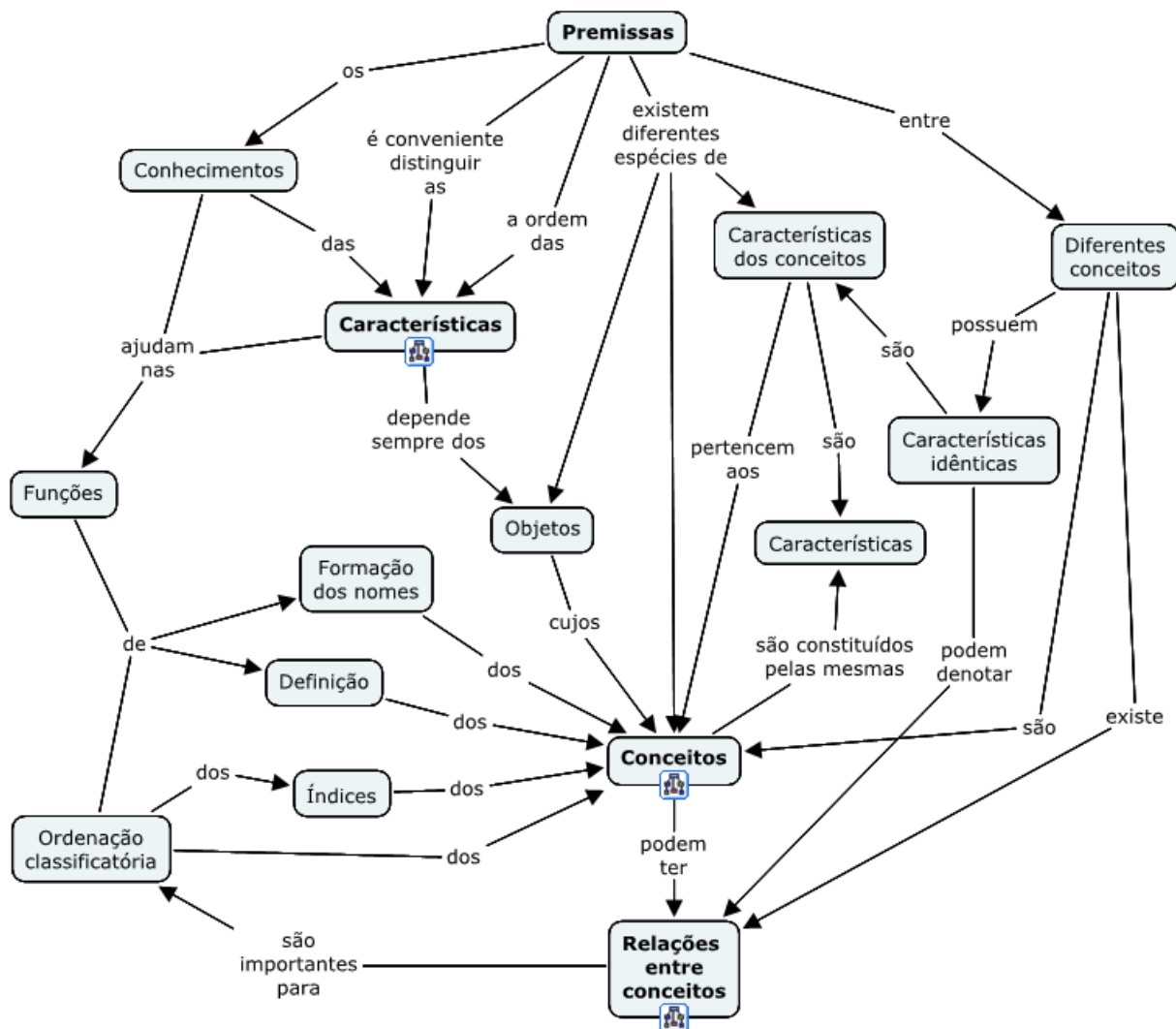
## Apêndice E - Mapa conceitual sobre a análise de intenção e extensão



### Apêndice F - Síntese das características de um conceito - análise intensional

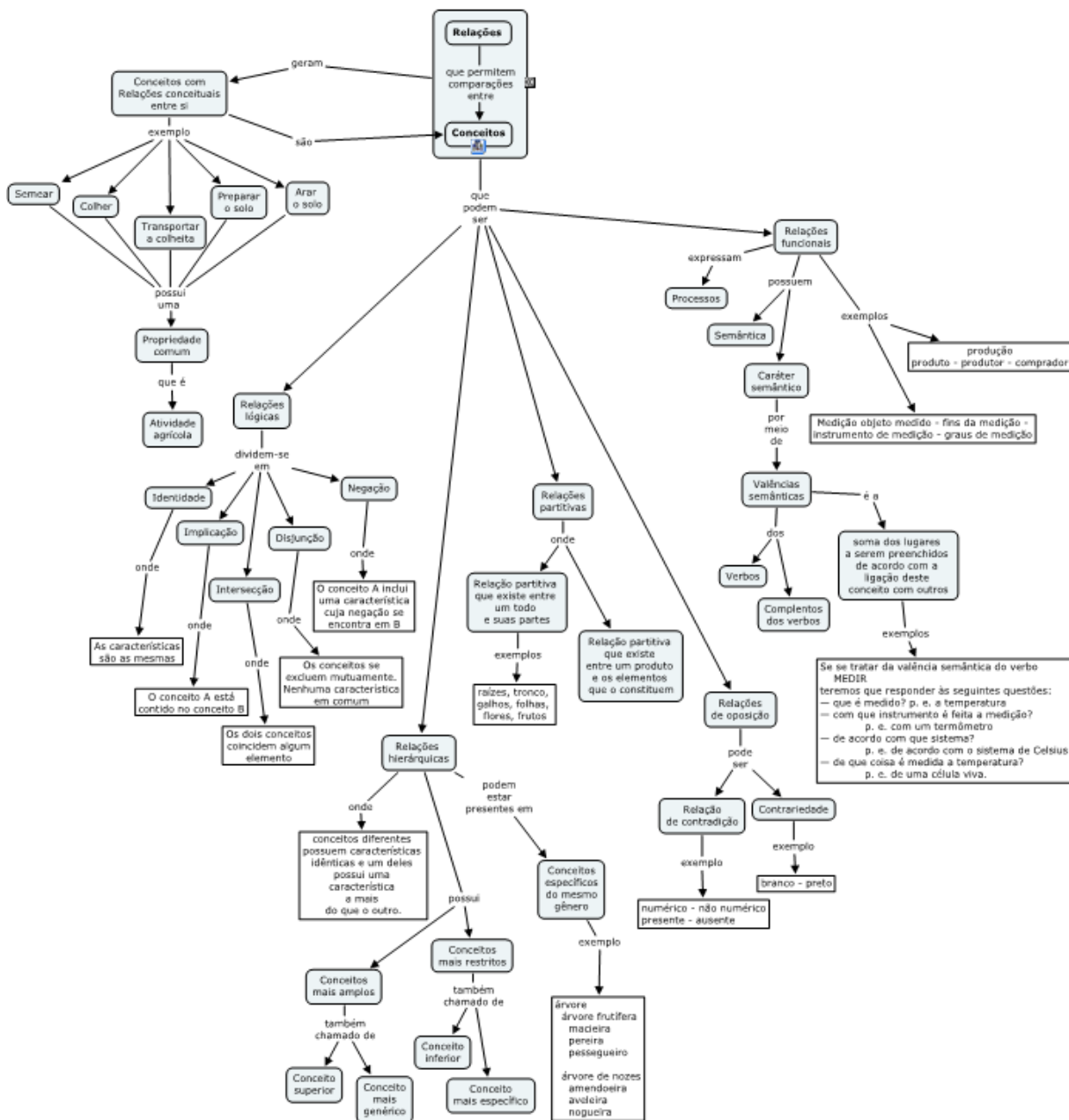


Apêndice G - Premissas para conceitualização – análise intensional

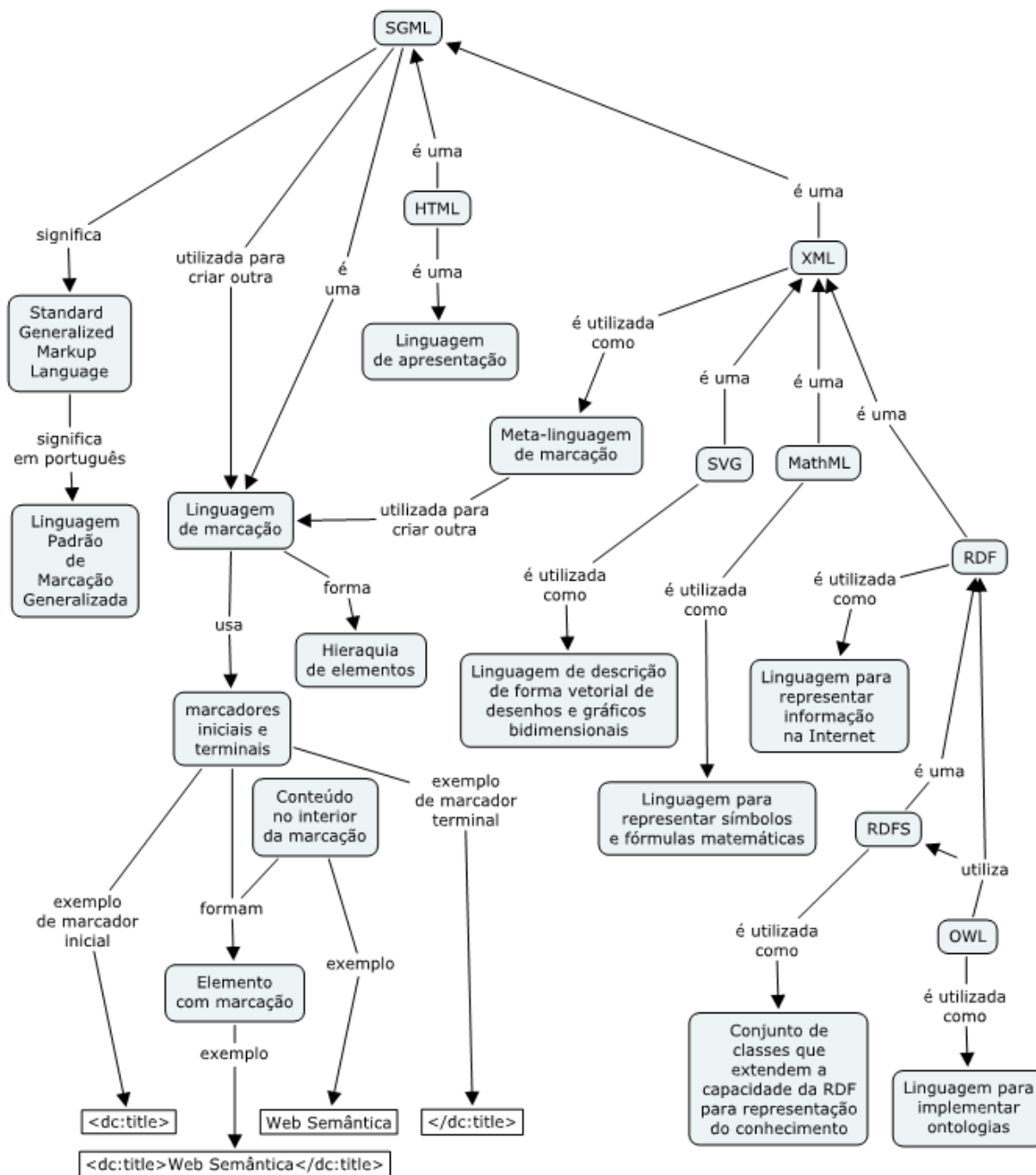




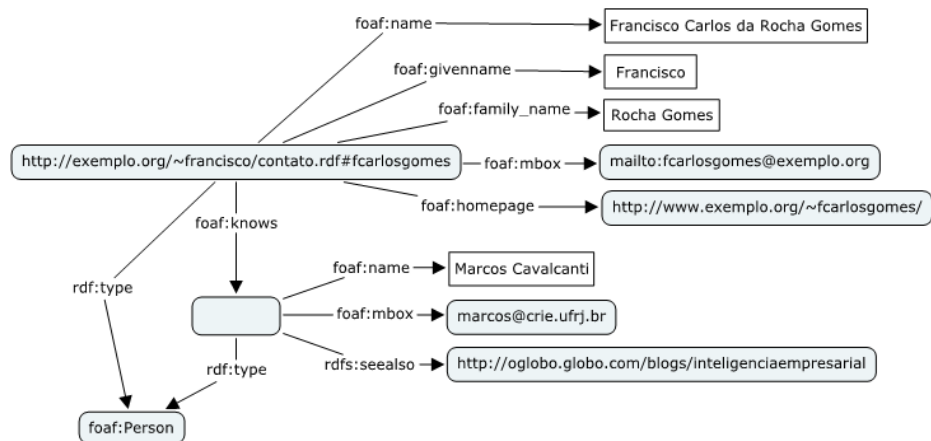
Apêndice H - Relações entre conceitos para análise extensional



Apêndice I - Mapa conceitual da linguagem genérica de marcação SGML



## Apêndice J - Exemplo de uso de vocabulários RDF e FOAF



```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <foaf:Person rdf:about="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">

    <foaf:name>Francisco Carlos da Rocha Gomes</foaf:name>
    <foaf:givenname>Francisco</foaf:givenname>
    <foaf:family_name>Rocha Gomes</foaf:family_name>
    <foaf:mbox rdf:resource="mailto:fcarlosgomes@exemplo.org"/>
    <foaf:homepage rdf:resource="http://www.exemplo.org/~fcarlosgomes/">

    <foaf:knows>
      <foaf:Person>
        <foaf:name>Marcos Cavalcanti</foaf:name>
        <foaf:mbox rdf:resource="marcos@crie.ufrj.br"/>
        <rdfs:seealso rdf:resource="http://oglobo.globo.com/blogs/inteligenciaempresarial"/>
      </foaf:Person>
    </foaf:knows>

  </foaf:Person>
</rdf:RDF>
```

Num	Sujeito	Predicado	Objeto
1	http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
2	http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes	http://xmlns.com/foaf/0.1/name	Francisco Carlos da Rocha Gomes
3	http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes	http://xmlns.com/foaf/0.1/givenname	Francisco
4	http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes	http://xmlns.com/foaf/0.1/family_name	Rocha Gomes
5	http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes	http://xmlns.com/foaf/0.1/mbox	mailto:fcarlosgomes@exemplo.org
6	http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes	http://xmlns.com/foaf/0.1/homepage	http://www.exemplo.org/~fcarlosgomes/
7	genid:A14576	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
8	http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes	http://xmlns.com/foaf/0.1/knows	genid:A14576
9	genid:A14576	http://xmlns.com/foaf/0.1/name	Marcos Cavalcanti
10	genid:A14576	http://xmlns.com/foaf/0.1/mbox	http://www.w3.org/RDF/Validator/run/marcos@crie.ufrj.br
11	genid:A14576	http://www.w3.org/2000/01/rdf-schema#seealso	http://oglobo.globo.com/blogs/inteligenciaempresarial

## Apêndice K - Códigos em RDF/XML

Os código abaixo foram validados no W3C *Validation Service* - <http://www.w3.org/RDF/Validator/>.

### Exemplo de sujeito – predicado – objeto

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ex="http://www.exemplo.com/conceitos#">
  <rdf:Description rdf:about="http://www.example.com/concepts#umSujeito">
    <ex:umPredicado
      rdf:resource="http://www.example.com/concepts#umObjeto"/>
  </rdf:Description>
</rdf:RDF>
```

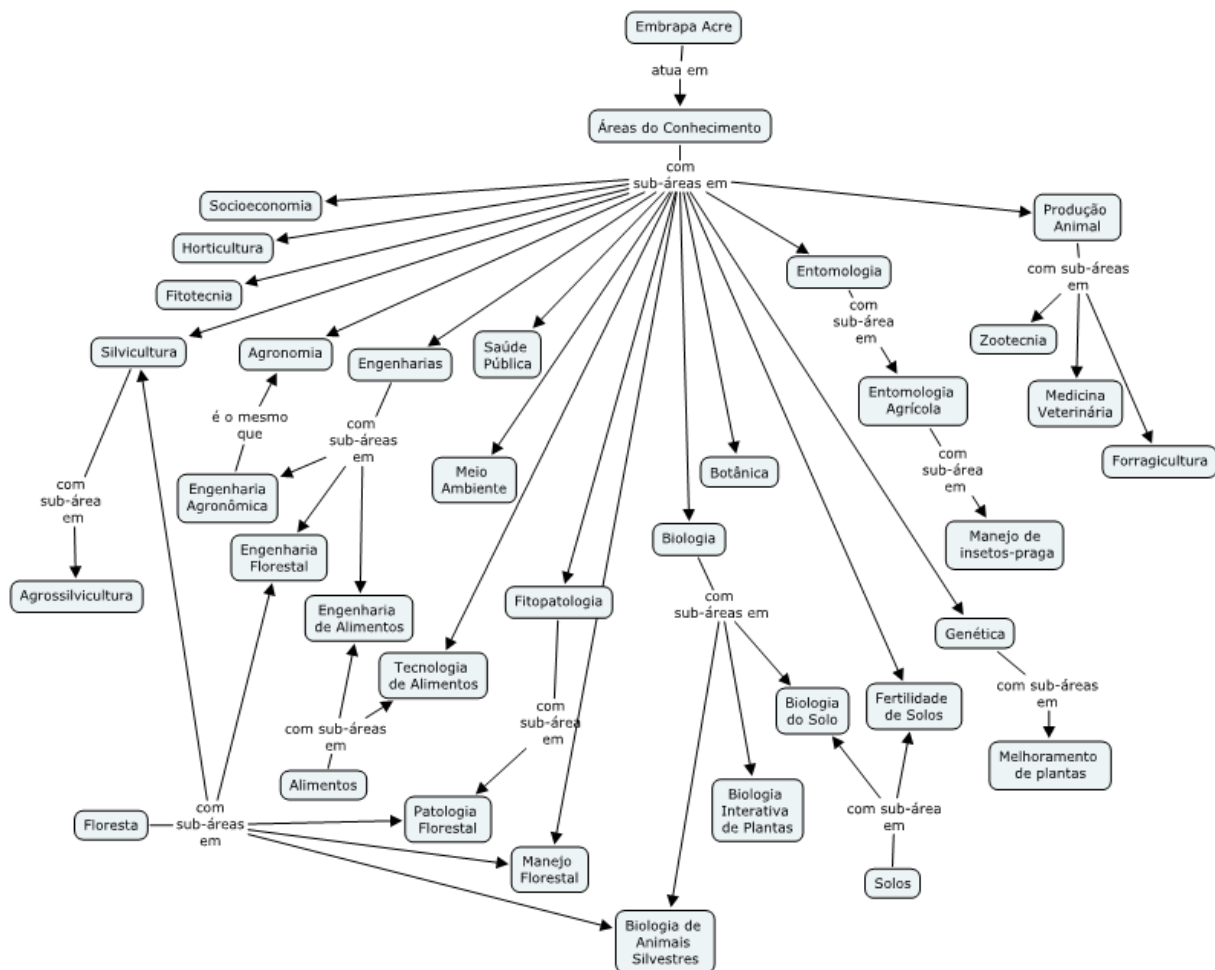
### Exemplo de definição de vocabulário com RDF/RDFS

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <rdf:Description rdf:about="vocabulario">
    <rdfs:comment>Este meu vocabulario descreve um dominio de conhecimento
  </rdfs:comment>
</rdf:Description>
  <rdfs:Class rdf:about="http://exemplo.org/vocabulario#Pessoa">
    <rdfs:isDefinedBy rdf:resource="http://exemplo.org/vocabulario#"/>
    <rdfs:label>Pessoa</rdfs:label>
    <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-
schema#Resource"/>
  </rdfs:Class>
  <rdfs:Property rdf:about="http://exemplo.org/vocabulario#nome">
    <rdfs:isDefinedBy rdf:resource="http://exemplo.org/vocabulario#"/>
  </rdfs:Property>
  <rdfs:Property rdf:about="http://exemplo.org/vocabulario#email">
    <rdfs:isDefinedBy rdf:resource="http://exemplo.org/vocabulario#"/>
  </rdfs:Property>
</rdf:RDF>
```

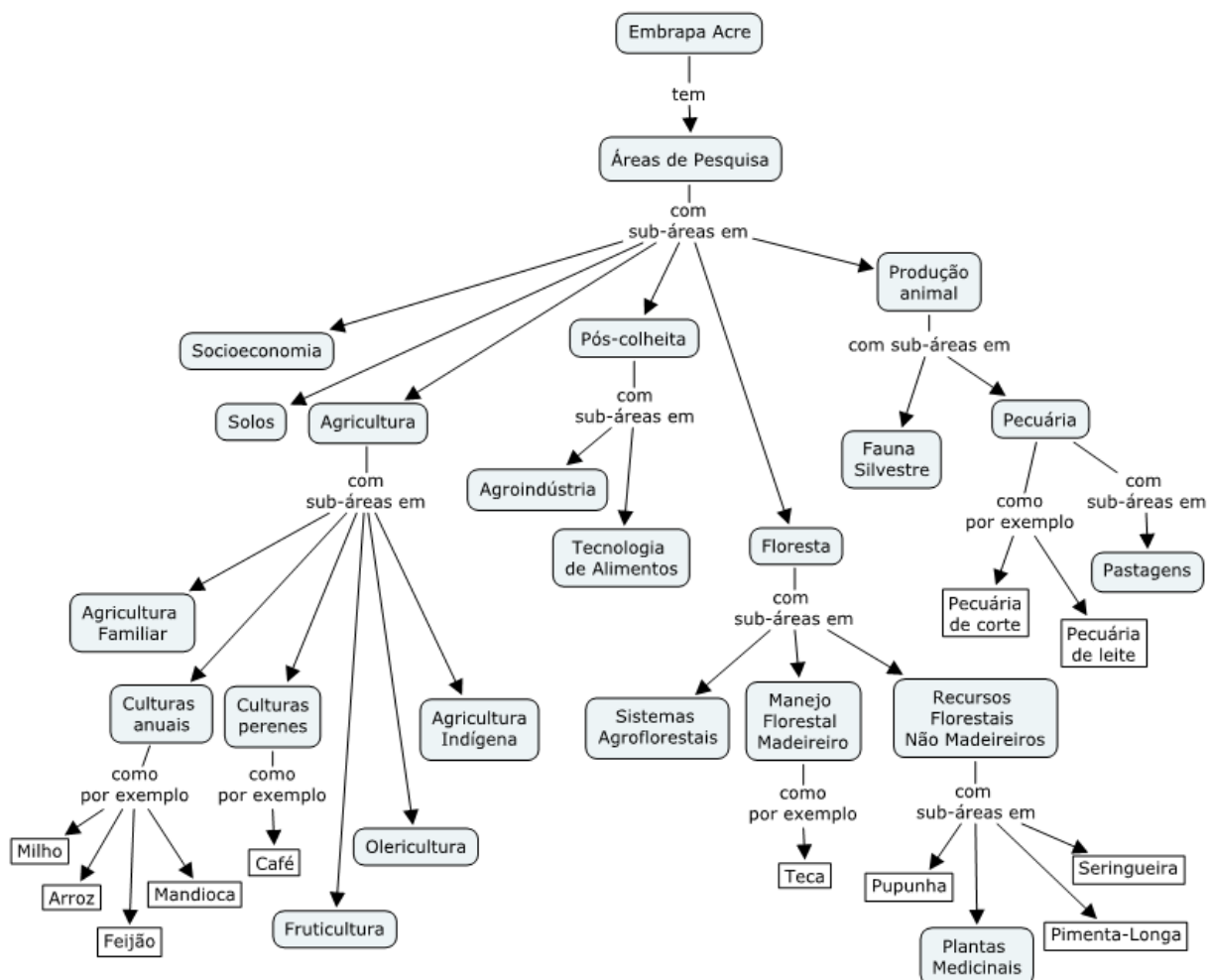
## Exemplo de especificação com vocabulários RDF e FOAF

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <foaf:Person
rdf:about="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">
  <foaf:name>Francisco Carlos da Rocha Gomes</foaf:name>
  <foaf:givenname>Francisco</foaf:givenname>
  <foaf:family_name>Rocha Gomes</foaf:family_name>
  <foaf:mbox rdf:resource="mailto:fcarlosgomes@exemplo.org"/>
  <foaf:homepage rdf:resource="http://www.exemplo.org/~fcarlosgomes"/>
  <foaf:knows>
    <foaf:Person>
      <foaf:name>Marcos Cavalcanti</foaf:name>
      <foaf:mbox rdf:resource="marcos@crie.ufrj.br"/>
      <rdfs:seealso
rdf:resource="http://oglobo.globo.com/blogs/inteligenciaempresarial"/>
      </foaf:Person>
    </foaf:knows>
  </foaf:Person>
</rdf:RDF>
```

Apêndice L - Áreas do conhecimento da Embrapa Acre



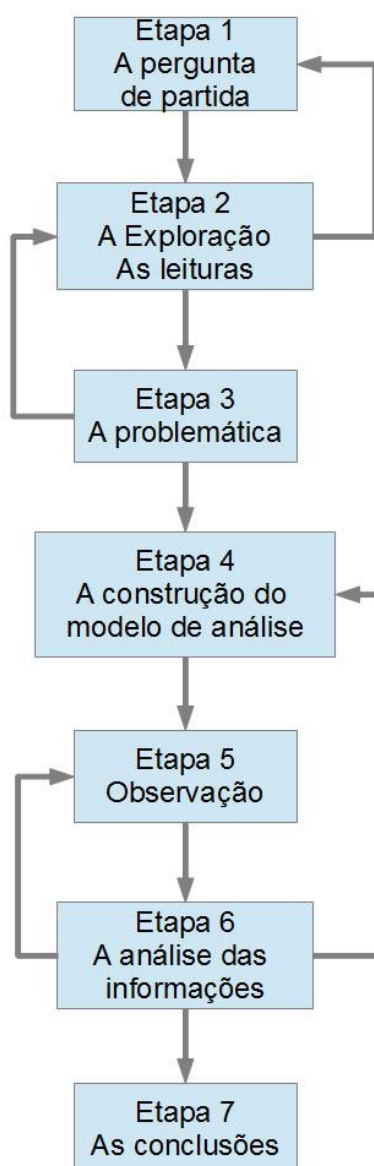
## Apêndice M - Áreas de Pesquisa da Embrapa Acre



## ANEXOS

### Anexo A - Esquema de procedimento científico

Esquema de procedimento científico definido por Quivy e Campenhoudt (1995, p. 27) para a condução de pesquisas em ciência sociais.





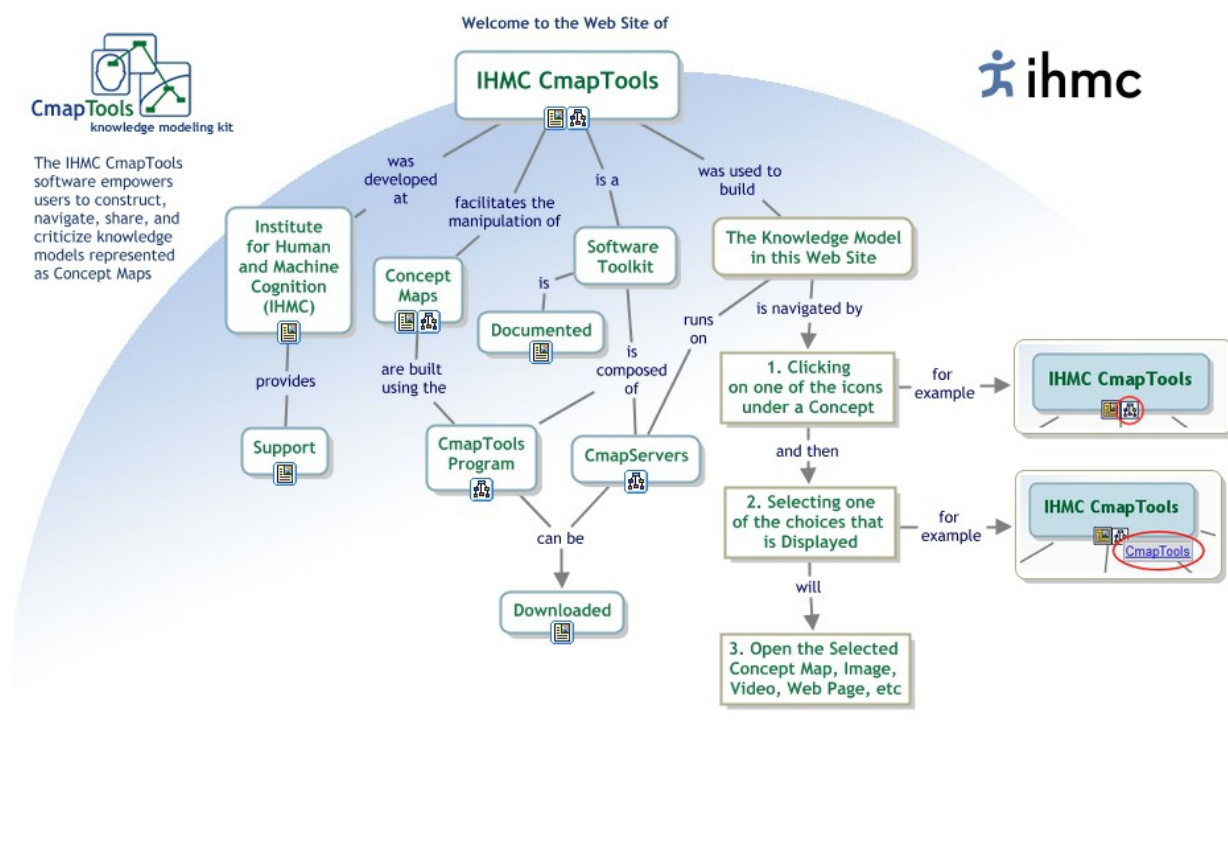
## Anexo B - Programa de Computador CmapTools

CmapTools é um programa de computador que auxilia na elaboração de mapas conceituais proporcionando rapidez e consistência ao processo.

É um *software* multiplataforma, utiliza a linguagem de programação Java, podendo ser instalado em diversos sistemas operacionais.

O programa de computador CmapTools pode ser obtido pelo site da Internet:  
<http://cmap.ihmc.us/>


Página principal na Internet do fornecedor do *software* CmapTools:




Página para *download* do software CmapTools:

**Download IHMC CmapTools**

If you would like to help keep CmapTools a free download — please consider making a donation (\$10 and up) by clicking one of the Donation buttons.





Windows	File Size	Notes
IHMC CmapTools v5.05.01 for Windows	84.9MB	512 MB of RAM / main memory or higher recommended
IHMC CmapTools v5.05.01 for Windows 64Bits	76MB	512 MB of RAM / main memory or higher recommended. Version for 64-bit architectures
IHMC CmapLite v5.05.01 for Windows	46.4MB	Recommended for smaller machines with a reduced memory size. CmapLite is also appropriate for older machines with small main memory. <a href="#">More details</a>

Mac OS X	File Size	Notes
IHMC CmapTools v5.05.01 for Mac OS X	55.6MB	Requires OS version 10.4 or higher, with latest Java VM. 512 MB of RAM / main memory or higher recommended
IHMC CmapTools v4.18 for Mac OS X	34MB	Recommended for OS version 10.3.9 (Panther) and previous versions
IHMC CmapLite v5.05.01 for Mac OS X	19.8MB	Requires OS version 10.4 or higher, with latest Java VM. Recommended for smaller machines with a reduced memory size. <a href="#">More details</a>

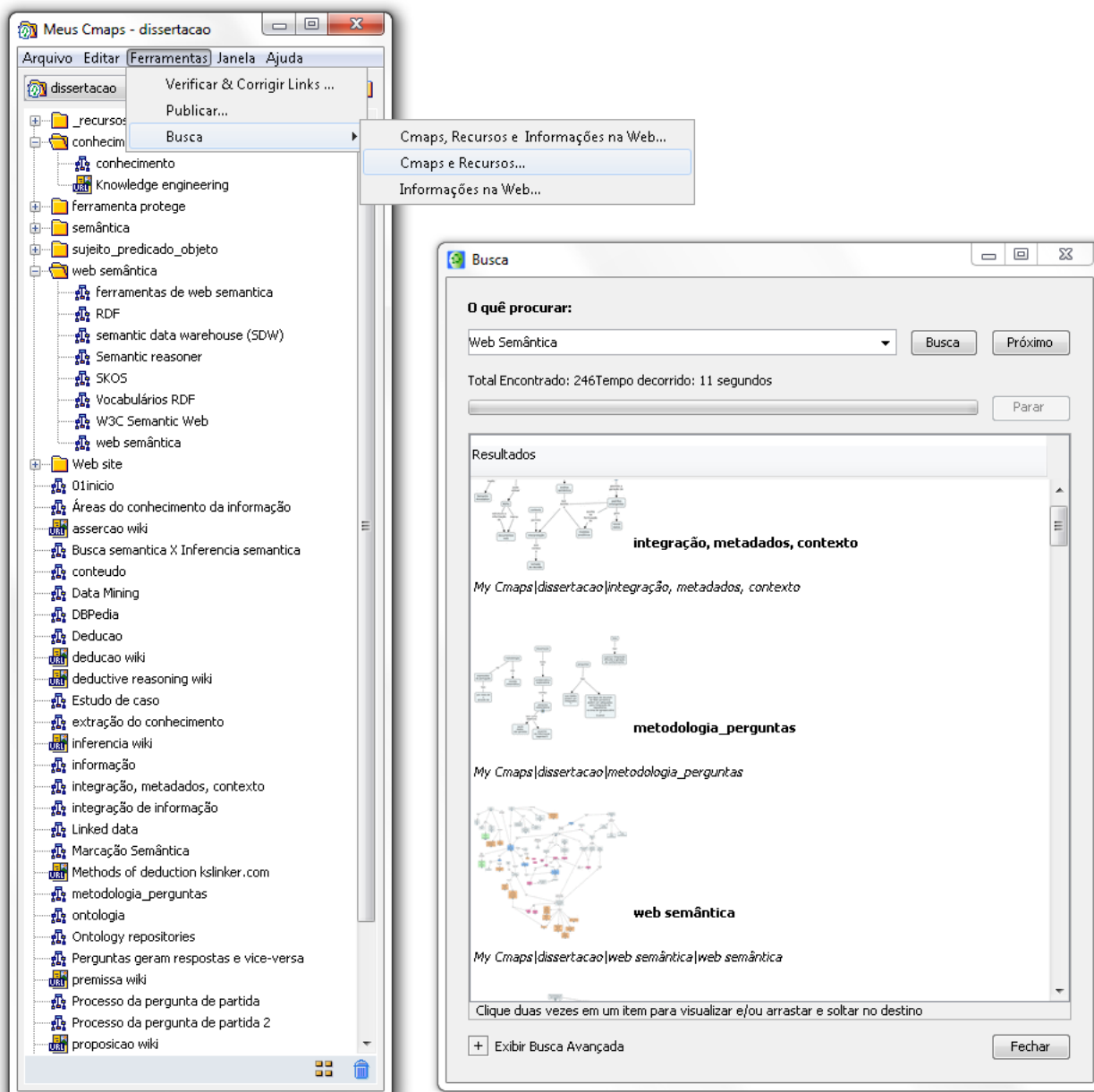
  

Linux	File Size	Notes
IHMC CmapTools v5.05.01 for Linux	166.9MB	512 MB of RAM / main memory or higher recommended
IHMC CmapTools v5.05.01 for Linux 64Bits	140NBMB	512 MB of RAM / main memory or higher recommended. Version for 64-bit architectures
IHMC CmapLite v5.05.01 for Linux	74MB	Recommended for smaller machines with a reduced memory size, in particular Netbooks. CmapLite is also appropriate for older machines with small main memory. <a href="#">More details</a>

Solaris	File Size	Notes
IHMC CmapTools v5.04.02 for Solaris	92.8MB	Requires SPARC System. 512 MB of RAM / main memory or higher recommended
IHMC CmapLite v5.04.01 for Solaris	64.6MB	Recommended to run on smaller machines with a reduced memory size. CmapLite is also appropriate for older machines with small main memory. <a href="#">More details</a>

Mecanismo de busca do *software* CmapTools empregado nesta dissertação para localização rápida de termos:



## Anexo C - Vocabulários controlados da web semântica

Prefix	Namespace	Title
bibtex	<a href="http://purl.org/net/nknouf/ns/bibtex#">http://purl.org/net/nknouf/ns/bibtex#</a>	BibTeX ontology
biol	<a href="http://purl.org/NET/biol/ns#">http://purl.org/NET/biol/ns#</a>	Biological Taxonomy Vocabulary 0.2 (Core)
botany	<a href="http://purl.org/NET/biol/botany#">http://purl.org/NET/biol/botany#</a>	Biological Taxonomy Vocabulary 0.2 (Botany)
dc	<a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a>	Dublin Core DCMI Metadata Terms
ep	<a href="http://eprints.org/ontology/">http://eprints.org/ontology/</a>	EPrints Ontology
foaf	<a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a>	Friend of a Friend vocabulary
food	<a href="http://data.lirmm.fr/ontologies/food#">http://data.lirmm.fr/ontologies/food#</a>	Food Ontology
skos	<a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#</a>	Simple Knowledge Organization System
food	<a href="http://www.w3.org/TR/2003/PR-owl-guide-20031209/food#">http://www.w3.org/TR/2003/PR-owl-guide-20031209/food#</a>	Food Ontology
gc	<a href="http://www.oegov.org/core/owl/gc#">http://www.oegov.org/core/owl/gc#</a>	oeGOV Government Core Ontology
gd	<a href="http://reference.data.gov/def/govdata/">http://reference.data.gov/def/govdata/</a>	Government Data Vocabulary
geo	<a href="http://www.w3.org/2003/01/geo/wgs84_pos#">http://www.w3.org/2003/01/geo/wgs84_pos#</a>	WGS84 Geo Positioning
geop	<a href="http://aims.fao.org/aos/geopolitical.owl#">http://aims.fao.org/aos/geopolitical.owl#</a>	FAO Geopolitical Ontology
cc	<a href="http://creativecommons.org/ns#">http://creativecommons.org/ns#</a>	Creative Commons Rights Expression Language
ceo	<a href="http://www.ebusiness-unibw.org/ontologies/consumerelectronics/v1#">http://www.ebusiness-unibw.org/ontologies/consumerelectronics/v1#</a>	Consumer Electronics Ontology
biotop	<a href="http://purl.org/biotop/biotop.owl#">http://purl.org/biotop/biotop.owl#</a>	BioTop
inno	<a href="http://purl.org/innovation/ns#">http://purl.org/innovation/ns#</a>	Ontology for Innovation
lgdo	<a href="http://linkedgeodata.org/ontology/">http://linkedgeodata.org/ontology/</a>	LinkedGeoData ontology
mo	<a href="http://purl.org/ontology/mo/">http://purl.org/ontology/mo/</a>	Music Ontology
music	<a href="http://www.kanzaki.com/ns/music#">http://www.kanzaki.com/ns/music#</a>	Music Vocabulary
oan	<a href="http://data.lirmm.fr/ontologies/oan/">http://data.lirmm.fr/ontologies/oan/</a>	Ontologie de l'Assemblée Nationale
odv	<a href="http://reference.data.gov.uk/def/organogram/">http://reference.data.gov.uk/def/organogram/</a>	Organogram Data Vocabulary
oecc	<a href="http://www.oegov.org/core/owl/cc#">http://www.oegov.org/core/owl/cc#</a>	Extended Creative Commons Ontology
rr	<a href="http://www.w3.org/ns/r2rml#">http://www.w3.org/ns/r2rml#</a>	RDB to RDF Mapping Language Schema
rss	<a href="http://purl.org/rss/1.0/">http://purl.org/rss/1.0/</a>	Vocabulary for Rich Site Summary (RSS) 1.0
scovo	<a href="http://purl.org/NET/scovo#">http://purl.org/NET/scovo#</a>	The Statistical Core Vocabulary
taxon	<a href="http://purl.org/biodiversity/taxon/">http://purl.org/biodiversity/taxon/</a>	TaxonMap Ontology
vcard	<a href="http://www.w3.org/2006/vcard/ns#">http://www.w3.org/2006/vcard/ns#</a>	An Ontology for vCards
tzont	<a href="http://www.w3.org/2006/timezone#">http://www.w3.org/2006/timezone#</a>	Time Zone Ontology
thea	<a href="http://purl.org/theatre#">http://purl.org/theatre#</a>	Theatre Ontology
time	<a href="http://www.w3.org/2006/time#">http://www.w3.org/2006/time#</a>	Time Ontology

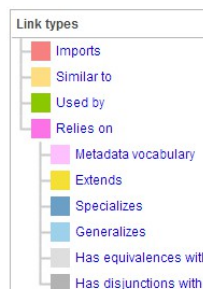
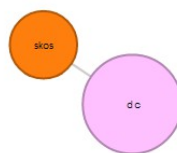
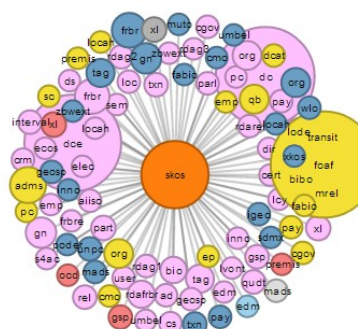
Fonte: Linked Open Vocabularies (LOV)  
<http://lov.okfn.org/dataset/lov/index.html>

## Anexo D - Referências e histórico do vocabulário SKOS

Vocabulary links:

Vocabularies referencing "skos" (93)

Vocabularies referenced by "skos" (1)



Vocabulary history:



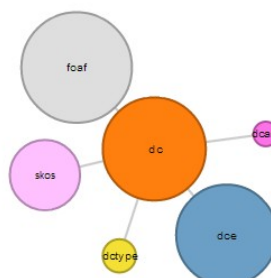
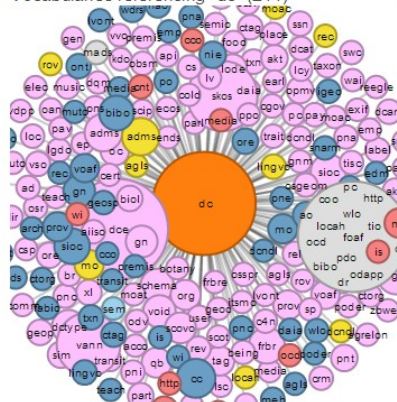
Fonte: Linked Open Vocabularies (LOV)  
[http://lov.okfn.org/dataset/lov/details/vocabulary\\_skos.html](http://lov.okfn.org/dataset/lov/details/vocabulary_skos.html)

## Referências e histórico do vocabulário Dublin Core (DC)

Vocabulary links:

Vocabularies referencing "dc" (211)

Vocabularies referenced by "dc" (5)



Vocabulary history:



Fonte: Linked Open Vocabularies (LOV)  
[http://lov.okfn.org/dataset/lov/details/vocabulary\\_dc.html](http://lov.okfn.org/dataset/lov/details/vocabulary_dc.html)

## Anexo D - Vocabulários SKOS

### Lista de vocabulários

← → ↻ 🏠 [www.w3.org/2001/sw/wiki/SKOS/Datasets](http://www.w3.org/2001/sw/wiki/SKOS/Datasets)

**W3C**  
Semantic Web

page discussion view source history

## SKOS/Datasets

< [SKOS](#)

**Contents** [\[hide\]](#)

**1 Find & Share SKOS Data**

- [1.1 STW Thesaurus for Economics](#)
- [1.2 Thesaurus for the Social Sciences](#)
- [1.3 Country Codes](#)
- [1.4 Drug Administration Forms](#)
- [1.5 Vocabularies for GeoSciML Geoscience information interchange](#)
- [1.6 DDC Dewey Decimal Classification](#)
- [1.7 Libris' vocabularies](#)
- [1.8 National Széchényi Library's vocabularies](#)
- [1.9 Library of Congress' vocabularies](#)
- [1.10 German national Library' subject headings](#)
- [1.11 French National Library's subject headings](#)
- [1.12 VIAF person authorities](#)
- [1.13 Wikipedia categories](#)
- [1.14 New York Times subjects](#)
- [1.15 IVOA astronomy vocabularies](#)
- [1.16 NASA taxonomy](#)
- [1.17 GEMET General Multilingual Environmental Thesaurus](#)
- [1.18 AGROVOC Agricultural Thesaurus](#)
- [1.19 Linked Life Data](#)
- [1.20 Taxonconcept](#)
- [1.21 UK Public sector vocabularies](#)
- [1.22 UMBEL Upper Mapping and Binding Exchange Layer](#)
- [1.23 Blogger's Topics](#)
- [1.24 GTAA, MeSH and IPSV](#)
- [1.25 UKAT UK Archival Thesaurus](#)
- [1.26 W3C Glossaries](#)
- [1.27 Language codes](#)
- [1.28 IPTC](#)
- [1.29 Government of Canada Core Subject Thesaurus](#)
- [1.30 Web NDL Authorities from National Diet Library of Japan](#)
- [1.31 Australian education vocabularies](#)
- [1.32 Geological Survey Austria](#)
- [1.33 reegle Clean Energy and Climate Change Thesaurus](#)
- [1.34 Social Semantic Web Thesaurus](#)
- [1.35 Cadastre and Land Administration Thesaurus](#)
- [1.36 Adequate, an ontology over peoples interests](#)
- [1.37 Proposed international standard nomenclature for fields of science and technology](#)
- [1.38 UNESCO Thesaurus](#)
- [1.39 Eurovoc](#)

**navigation**

- [Main Page](#)
- [Recent changes](#)
- [Tools](#)
- [Books](#)
- [Validators](#)

**other w3c resources:**

- [Activity news](#)
- [Publications](#)
- [Logos, buttons](#)
- [Activity home page](#)

**w3c rss feeds**

- [Activity newsfeed](#)
- [W3C blogs](#)
- [Use cases, case studies](#)

**account request**

- [W3C Member](#)
- [Public](#)

**search**

**toolbox**

- [What links here](#)
- [Related changes](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)
- [Browse properties](#)

Fonte: Conjunto de vocabulários SKOS  
<http://www.w3.org/2001/sw/wiki/SKOS/Datasets>

## Tesauro AGROVOC da FAO

← → ↻ 🏠 aims.fao.org/advice/linked-data

english français español 中文

**AIMS** Agricultural Information Management Standards

Username: \* Password: \*  
   
 LOG IN > request new password

Food and Agriculture Organization of the United Nations  
 for a world without hunger

Home » Linked Data  
**Linked Data**

Getting Started AGROVOC LOD LODE-BD OpenAGRIS

A number of initiatives related to the Semantic Web are taking place within the AIMS umbrella. These innovative experiences are meant to support those institutions that wish to consume and publish Linked Data as a means of sharing and exchange agricultural research information in the context of new challenges to agriculture. These initiatives are:

**Vocabularies published as Linked Data**

AGROVOC which is linked to more than a dozen vocabularies

**Recommendations to select appropriate encoding strategies for producing bibliographic data suitable for the Linked Data environment**

The Linked Open Data - Enabled Bibliographic Data (LODE-BD)  
 Recommendations that provides encoding strategies for producing LOD-enabled bibliographical data as well as the encoding of value vocabularies used in describing agents, places, and topics in bibliographic data

**Web applications to aggregate information from different Web sources to existing data repositories**

OpenAGRIS beta version works expanding the AGRIS repository interlinking to other datasets.

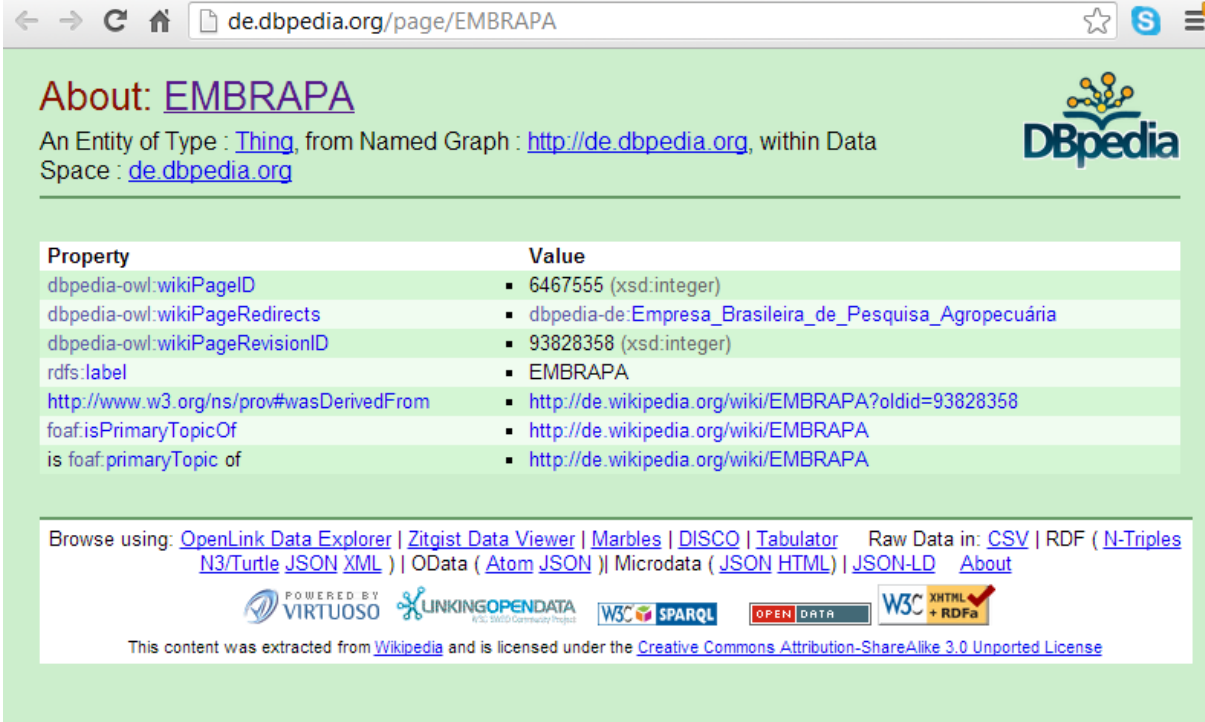
A glossary of terms related to Linked Data and as well external resources are available from the get started section.

VOTE 28 🍌 SHARE Tweet 1 +

Fonte: AIMS – Agricultural Information Management Standards  
<http://aims.fao.org/advice/linked-data>

## Anexo E - Termos da DBpedia

### Embrapa



The screenshot shows a web browser window displaying the DBpedia page for 'Embrapa'. The address bar shows 'de.dbpedia.org/page/EMBRAPA'. The page title is 'About: EMBRAPA'. Below the title, it states: 'An Entity of Type : [Thing](#), from Named Graph : <http://de.dbpedia.org>, within Data Space : <de.dbpedia.org>'. The DBpedia logo is visible in the top right corner.

Property	Value
dbpedia-owl:wikiPageID	▪ 6467555 (xsd:integer)
dbpedia-owl:wikiPageRedirects	▪ dbpedia-de:Empresa_Brasileira_de_Pesquisa_Agropecuária
dbpedia-owl:wikiPageRevisionID	▪ 93828358 (xsd:integer)
rdfs:label	▪ EMBRAPA
<a href="http://www.w3.org/ns/prov#wasDerivedFrom">http://www.w3.org/ns/prov#wasDerivedFrom</a>	▪ <a href="http://de.wikipedia.org/wiki/EMBRAPA?oldid=93828358">http://de.wikipedia.org/wiki/EMBRAPA?oldid=93828358</a>
foaf:isPrimaryTopicOf	▪ <a href="http://de.wikipedia.org/wiki/EMBRAPA">http://de.wikipedia.org/wiki/EMBRAPA</a>
is foaf:primaryTopic of	▪ <a href="http://de.wikipedia.org/wiki/EMBRAPA">http://de.wikipedia.org/wiki/EMBRAPA</a>

Below the table, there are links for browsing the data: 'Browse using: [OpenLink Data Explorer](#) | [Zitgist Data Viewer](#) | [Marbles](#) | [DISCO](#) | [Tabulator](#) Raw Data in: [CSV](#) | [RDF \( N-Triples N3/Turtle JSON XML \)](#) | [OData \( Atom JSON \)](#) | [Microdata \( JSON HTML \)](#) | [JSON-LD](#) [About](#)

At the bottom, there are logos for 'POWERED BY VIRTUOSO', 'LINKINGOPENDATA', 'W3C SPARQL', 'OPEN DATA', and 'W3C XHTML + RDFa'. A footer note states: 'This content was extracted from [Wikipedia](#) and is licensed under the [Creative Commons Attribution-ShareAlike 3.0 Unported License](#)'.

Fonte: DBpedia  
<http://de.dbpedia.org/page/EMBRAPA>




## Cereal

← → ↻ ↑ [live.dbpedia.org/page/Cereal](http://live.dbpedia.org/page/Cereal) ☆ S

### About: Cereal

An Entity of Type : [plant](#), from Named Graph : <http://live.dbpedia.org>, within Data Space : [live.dbpedia.org](http://live.dbpedia.org)



A cereal is a grass cultivated for the edible components of their grain (botanically, a type of fruit called a caryopsis), composed of the endosperm, germ, and bran. Cereal grains are grown in greater quantities and provide more food energy worldwide than any other type of crop; they are therefore staple crops. In their natural form, they are a rich source of vitamins, minerals, carbohydrates, fats, oils, and protein.

Property	Value
<a href="#">dbpedia-owl:abstract</a>	<ul style="list-style-type: none"> <li>A cereal is a grass cultivated for the edible components of their grain (botanically, a type of fruit called a caryopsis), composed of the endosperm, germ, and bran. Cereal grains are grown in greater quantities and provide more food energy worldwide than any other type of crop; they are therefore staple crops. In their natural form, they are a rich source of vitamins, minerals, carbohydrates, fats, oils, and protein. However, when refined and processed, the remaining endosperm is mostly carbohydrate and lacks the majority of the nutrients. In developing nations, grain in the form of rice, wheat, millet, or maize constitutes a major part of the diet. In developed nations, cereal consumption is moderate and varied but still substantial. The goddess Ceres, the name of the Roman goddess of harvest and agriculture.</li> </ul>
<a href="#">dbpedia-owl:class</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Monocots</a></li> </ul>
<a href="#">dbpedia-owl:division</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Angiosperms</a></li> </ul>
<a href="#">dbpedia-owl:family</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Poaceae</a></li> </ul>
<a href="#">dbpedia-owl:kingdom</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Plantae</a></li> </ul>
<a href="#">dbpedia-owl:order</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Commelinids</a></li> <li><a href="#">dbpedia:Poales</a></li> </ul>
<a href="#">dbpedia-owl:thumbnail</a>	<ul style="list-style-type: none"> <li><a href="http://upload.wikimedia.org/wikipedia/commons/thumb/1/17/Various_grains_edit2.jpg/240px-Variou...">http://upload.wikimedia.org/wikipedia/commons/thumb/1/17/Various_grains_edit2.jpg/240px-Variou...</a></li> </ul>
<a href="#">dbpedia-owl:wikiPageExternalLink</a>	<ul style="list-style-type: none"> <li><a href="http://www.vegsoc.org/info/cereals.html">http://www.vegsoc.org/info/cereals.html</a></li> <li><a href="http://books.nap.edu/catalog.php?record_id=2305">http://books.nap.edu/catalog.php?record_id=2305</a></li> <li><a href="http://www.hgca.com">http://www.hgca.com</a></li> <li><a href="http://www.cereal.com/recipes.htm">http://www.cereal.com/recipes.htm</a></li> </ul>
<a href="#">dbpprop: familia</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Poaceae</a></li> </ul>
<a href="#">dbpprop: imageCaption</a>	<ul style="list-style-type: none"> <li>Various cereals and their products</li> </ul>
<a href="#">dbpprop: imageWidth</a>	<ul style="list-style-type: none"> <li>250 (xsd:integer)</li> </ul>
<a href="#">dbpprop: ordo</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Poales</a></li> </ul>
<a href="#">dbpprop: regnum</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Plantae</a></li> </ul>
<a href="#">dbpprop: unrankedClassis</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Monocots</a></li> </ul>
<a href="#">dbpprop: unrankedDivisio</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Angiosperms</a></li> </ul>
<a href="#">dbpprop: unrankedOrdo</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Commelinids</a></li> </ul>
<a href="#">dbpprop: wikiPageUsesTemplate</a>	<ul style="list-style-type: none"> <li><a href="#">dbpedia:Template:Taxobox</a></li> </ul>
<a href="#">dcterms: subject</a>	<ul style="list-style-type: none"> <li><a href="#">category:Crops</a></li> <li><a href="#">category:Cereals</a></li> </ul>
<a href="#">rdf: type</a>	<ul style="list-style-type: none"> <li><a href="#">owl:Thing</a></li> <li><a href="#">dbpedia-owl:Eukaryote</a></li> <li><a href="#">dbpedia-owl:Plant</a></li> <li><a href="#">dbpedia-owl:Species</a></li> </ul>
<a href="#">rdfs: comment</a>	<ul style="list-style-type: none"> <li>A cereal is a grass cultivated for the edible components of their grain (botanically, a type of fruit called a caryopsis), composed of the endosperm, germ, and bran. Cereal grains are grown in greater quantities and provide more food energy worldwide than any other type of crop; they are therefore staple crops. In their natural form, they are a rich source of vitamins, minerals, carbohydrates, fats, oils, and protein.</li> </ul>
<a href="#">rdfs: label</a>	<ul style="list-style-type: none"> <li>Cereal</li> </ul>

Fonte: DBPedia

<http://live.dbpedia.org/page/Cereal>

## Cassava (Mandioca)

← → ↻ 🏠 📄 live.dbpedia.org/page/Cassava ☆ S ☰

### About: Cassava

An Entity of Type : [plant](#), from Named Graph : <http://live.dbpedia.org>, within Data Space : [live.dbpedia.org](http://live.dbpedia.org)



Cassava (*Manihot esculenta*), also called yuca, mogo, manioc, mandioca, and kamoteng kahoy, a woody shrub of the Euphorbiaceae (spurge family) native to South America, is extensively cultivated as an annual crop in tropical and subtropical regions for its edible starchy, tuberous root, a major source of carbohydrates. It differs from the similarly-spelled yucca, an unrelated fruit-bearing shrub in the Asparagaceae family.

Property	Value
dbpedia-owl:abstract	<ul style="list-style-type: none"> <li>Cassava (<i>Manihot esculenta</i>), also called yuca, mogo, manioc, mandioca, and kamoteng kahoy, is extensively cultivated as an annual crop in tropical and subtropical regions for its edible starchy tuberous root, a major source of carbohydrates. It differs from the similarly-spelled yucca, an unrelated fruit-bearing shrub in the Asparagaceae family. Cassava is the third-largest source of food carbohydrates in the tropics. Cassava people. It is one of the most drought-tolerant crops, capable of growing on marginal soils. Nig carbohydrates, but a poor source of protein. A predominantly cassava root diet can cause pr tubers. Cassava contains anti-nutrition factors and toxins. It must be properly prepared before cause acute cyanide intoxication and goiters, and may even cause ataxia or partial paralysis and thieves. The more-toxic varieties of cassava are a fall-back resource (a "food security cro</li> </ul>
dbpedia-owl:class	<ul style="list-style-type: none"> <li>dbpedia:Eudicots</li> </ul>
dbpedia-owl:division	<ul style="list-style-type: none"> <li>dbpedia:Angiosperms</li> </ul>
dbpedia-owl:family	<ul style="list-style-type: none"> <li>dbpedia:Crotonoideae</li> <li>dbpedia:Manihoteae</li> <li>dbpedia:Euphorbiaceae</li> </ul>
dbpedia-owl:genus	<ul style="list-style-type: none"> <li>dbpedia:Manihot</li> </ul>
dbpedia-owl:kingdom	<ul style="list-style-type: none"> <li>dbpedia:Plant</li> </ul>
dbpedia-owl:order	<ul style="list-style-type: none"> <li>dbpedia:Rosids</li> <li>dbpedia:Malpighiales</li> </ul>
dbpedia-owl:thumbnail	<ul style="list-style-type: none"> <li><a href="http://upload.wikimedia.org/wikipedia/commons/thumb/f/f1/Manihot_esculenta_-_Köhler-s_M">http://upload.wikimedia.org/wikipedia/commons/thumb/f/f1/Manihot_esculenta_-_Köhler-s_M</a></li> </ul>
dbpedia-owl:wikiPageExternalLink	<ul style="list-style-type: none"> <li><a href="http://researchnews.osu.edu/archive/suprtubr.htm">http://researchnews.osu.edu/archive/suprtubr.htm</a></li> <li><a href="http://www.cassavabiz.org/">http://www.cassavabiz.org/</a></li> <li><a href="http://www.scielo.br/scielo.php?script=sci_arttext&amp;pid=S0104-79301996000100002&amp;lng=en8">http://www.scielo.br/scielo.php?script=sci_arttext&amp;pid=S0104-79301996000100002&amp;lng=en8</a></li> <li><a href="http://www.fao.org/ag/AGP/AGPC/gcds/GCS.htm">http://www.fao.org/ag/AGP/AGPC/gcds/GCS.htm</a></li> <li><a href="http://researchnews.osu.edu/archive/cassava.htm">http://researchnews.osu.edu/archive/cassava.htm</a></li> <li><a href="http://www.iita.org/cms/details/cassava_project_details.aspx?zoneid=63&amp;articleid=267">http://www.iita.org/cms/details/cassava_project_details.aspx?zoneid=63&amp;articleid=267</a></li> <li><a href="http://www.unu.edu/unupress/food/8F024e/8F024E01.htm">http://www.unu.edu/unupress/food/8F024e/8F024E01.htm</a></li> <li><a href="http://www.inoculatedmind.com/?p=121">http://www.inoculatedmind.com/?p=121</a></li> <li><a href="http://www.new-agri.co.uk/98-6/focuson/focuson9.html">http://www.new-agri.co.uk/98-6/focuson/focuson9.html</a></li> <li><a href="http://www.hort.purdue.edu/newcrop/CropFactSheets/cassava.html">http://www.hort.purdue.edu/newcrop/CropFactSheets/cassava.html</a></li> <li><a href="http://www.catisa.org/">http://www.catisa.org/</a></li> <li><a href="http://www.sciencedaily.com/releases/2009/06/090616133940.htm">http://www.sciencedaily.com/releases/2009/06/090616133940.htm</a></li> <li><a href="http://books.google.com.br/books/about/History_of_a_voyage_to_the_land_of_Brazi.html?id=">http://books.google.com.br/books/about/History_of_a_voyage_to_the_land_of_Brazi.html?id=</a></li> <li><a href="http://www.filipino-food-recipes.com/cassavarecipefilipino.html">http://www.filipino-food-recipes.com/cassavarecipefilipino.html</a></li> <li><a href="http://www.viruspadu.com/2008/10/manok-pansuh.html">http://www.viruspadu.com/2008/10/manok-pansuh.html</a></li> <li><a href="http://wendyinkk.blogspot.com/2012/09/manok-pansoh-sarawak-mff-1.html">http://wendyinkk.blogspot.com/2012/09/manok-pansoh-sarawak-mff-1.html</a></li> <li><a href="http://genuardis.net/pansoh/pansoh-ayam.htm">http://genuardis.net/pansoh/pansoh-ayam.htm</a></li> </ul>
dbpprop:binomial	<ul style="list-style-type: none"> <li>Manihot esculenta</li> </ul>
dbpprop:binomialAuthority	<ul style="list-style-type: none"> <li>Crantz</li> </ul>
dbpprop: familia	<ul style="list-style-type: none"> <li>dbpedia:Euphorbiaceae</li> </ul>
dbpprop:genus	<ul style="list-style-type: none"> <li>Manihot</li> </ul>
dbpedia: image	<ul style="list-style-type: none"> <li>Manihot esculenta 001.jpg</li> </ul>

Fonte: DBPedia  
<http://live.dbpedia.org/page/Cassava>

## Anexo F - Validação de RDF

### Tela inicial de validação

← → ↻ 🏠  ☆

**W3C** **RDF** **Powered** **Validation Service**

Home Documentation Feedback Donate

### Check and Visualize your RDF documents

Enter a URI or paste an RDF/XML document into the text field above. A 3-tuple (triple) representation of the corresponding data model as well as an optional graphical visualization of the data model will be displayed.

**Check by Direct Input**

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns="http://www.exemplo.org/~francisco/contato.rdf#">

  <foaf:Person
    rdf:about="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes"
  >
```

Parse RDF Restore the original example Clear the textarea

**Display Result Options:**  
 Triples and/or Graph: Triples Only  
 Graph format: PNG - embedded

Paste an RDF/XML document into the following text field to have it checked. More options are available in the [Extended interface](#).

**Check by URI**

Parse URI: Clear the URI

**Display Result Options:**  
 Triples and/or Graph: Triples Only  
 Graph format: PNG - embedded

Enter the URI for the RDF/XML document you would like to check. More options are available in the [Extended interface](#)

Fonte: W3C Validation Service  
<http://www.w3.org/RDF/Validator/>

## Resultado da Validação de RDF

www.w3.org/RDF/Validator/ARPServlet

**W3C** **RDF** **Powered** **Validation Service**

Home Documentation Feedback

### Validation Results

Your RDF document validated successfully.

#### Triples of the Data Model

Number	Subject	Predicate	Object
1	<a href="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes</a>	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	<a href="http://xmlns.com/foaf/0.1/Person">http://xmlns.com/foaf/0.1/Person</a>
2	<a href="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes</a>	<a href="http://xmlns.com/foaf/0.1/name">http://xmlns.com/foaf/0.1/name</a>	"Francisco Carlos da Rocha Gomes"
3	<a href="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes</a>	<a href="http://xmlns.com/foaf/0.1/givenname">http://xmlns.com/foaf/0.1/givenname</a>	"Francisco"
4	<a href="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes</a>	<a href="http://xmlns.com/foaf/0.1/family_name">http://xmlns.com/foaf/0.1/family_name</a>	"Rocha Gomes"
5	<a href="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes</a>	<a href="http://xmlns.com/foaf/0.1/mbox">http://xmlns.com/foaf/0.1/mbox</a>	<a href="mailto:fcarlosgomes@exemplo.org">mailto:fcarlosgomes@exemplo.org</a>
6	<a href="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes</a>	<a href="http://xmlns.com/foaf/0.1/homepage">http://xmlns.com/foaf/0.1/homepage</a>	<a href="http://www.exemplo.org/~fcarlosgomes/">http://www.exemplo.org/~fcarlosgomes/</a>
7	genid:A14129	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	<a href="http://xmlns.com/foaf/0.1/Person">http://xmlns.com/foaf/0.1/Person</a>
8	<a href="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes</a>	<a href="http://xmlns.com/foaf/0.1/knows">http://xmlns.com/foaf/0.1/knows</a>	genid:A14129
9	genid:A14129	<a href="http://xmlns.com/foaf/0.1/name">http://xmlns.com/foaf/0.1/name</a>	"Marcos Cavalcanti"
10	genid:A14129	<a href="http://xmlns.com/foaf/0.1/mbox">http://xmlns.com/foaf/0.1/mbox</a>	<a href="http://www.w3.org/RDF/Validator/run/marcos@crie.ufrj.br">http://www.w3.org/RDF/Validator/run/marcos@crie.ufrj.br</a>
11	genid:A14129	<a href="http://www.w3.org/2000/01/rdf-schema#seealso">http://www.w3.org/2000/01/rdf-schema#seealso</a>	<a href="http://oglobo.globo.com/blogs/inteligenciaempresarial">http://oglobo.globo.com/blogs/inteligenciaempresarial</a>

#### The original RDF/XML document

```
1: <rdf:RDF
2:   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3:   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
4:   xmlns:foaf="http://xmlns.com/foaf/0.1/"
5:   xmlns="http://www.exemplo.org/~francisco/contato.rdf#">
6:
7:   <foaf:Person rdf:about="http://www.exemplo.org/~francisco/contato.rdf#fcarlosgomes">
8:
9:     <foaf:name>Francisco Carlos da Rocha Gomes</foaf:name>
10:    <foaf:givenname>Francisco</foaf:givenname>
11:    <foaf:family_name>Rocha Gomes</foaf:family_name>
12:    <foaf:mbox rdf:resource="mailto:fcarlosgomes@exemplo.org"/>
13:    <foaf:homepage rdf:resource="http://www.exemplo.org/~fcarlosgomes/">
14:
15:    <foaf:knows>
16:      <foaf:Person>
17:        <foaf:name>Marcos Cavalcanti</foaf:name>
18:        <foaf:mbox rdf:resource="marcos@crie.ufrj.br"/>
19:        <rdfs:seealso rdf:resource="http://oglobo.globo.com/blogs/inteligenciaempresarial"/>
20:      </foaf:Person>
21:    </foaf:knows>
22:
23:  </foaf:Person>
24: </rdf:RDF>
25:
```

Fonte: W3C Validation Service  
<http://www.w3.org/RDF/Validator/>

## Anexo G - Organograma da Embrapa



Fonte: Embrapa

[http://www.embrapa.br/a\\_embrapa/Organograma-Embrapa](http://www.embrapa.br/a_embrapa/Organograma-Embrapa)

## Anexo H - Unidades da Embrapa no território nacional



Fonte: Embrapa

[http://hotsites.sct.embrapa.br/acessoainformacao/institucional/mapa1\\_uni\\_br.jpg](http://hotsites.sct.embrapa.br/acessoainformacao/institucional/mapa1_uni_br.jpg)

## Anexo I - Conexão do tesauro AgroVoc com a DBpedia

Tela do AgroVoc referente ao termo “maize” (milho). Em destaque o *link* para a página da DBpedia sobre milho.

The screenshot shows the AGROVOC Thesaurus page for the term 'maize'. The page displays a table of properties and their corresponding values. The 'closeMatch' property is highlighted with a red circle, indicating the link to the DBpedia resource for Maize.

Property	Value
altLabel	<ul style="list-style-type: none"> <li>Granoturco (it)</li> <li>Ziarno kukurydzy (pl)</li> <li>corn (maize) (en)</li> <li>kukorica (hu)</li> <li>zrno (kukurica) (sk)</li> <li>zrno kukuřice (cs)</li> <li>зерно кукурузы (ru)</li> <li>अनाज (मक्का) (hi)</li> <li>コーン (ja)</li> <li>苞谷 (zh)</li> </ul>
broader	<ul style="list-style-type: none"> <li>&lt;http://aims.fao.org/aos/agrovoc/c_1474&gt;</li> </ul>
closeMatch	<ul style="list-style-type: none"> <li>&lt;http://dbpedia.org/resource/Maize&gt;</li> </ul>
created	2011-11-20T20:35:16Z (dateTime)
exactMatch	<ul style="list-style-type: none"> <li>&lt;http://d-nb.info/gnd/4037135-9&gt;</li> <li>&lt;http://eurovoc.europa.eu/1744&gt;</li> <li>&lt;http://zbw.eu/stw/descriptor/14093-4&gt;</li> </ul>
hasStatus	Published (string)
isProducedBy	<http://aims.fao.org/aos/agrovoc/c_8504>
isUsedToMake	<ul style="list-style-type: none"> <li>&lt;http://aims.fao.org/aos/agrovoc/c_15500&gt;</li> <li>&lt;http://aims.fao.org/aos/agrovoc/c_25472&gt;</li> </ul>
modified	2012-06-26T22:11:59Z (dateTime)
narrower	<ul style="list-style-type: none"> <li>&lt;http://aims.fao.org/aos/agrovoc/c_14385&gt;</li> <li>&lt;http://aims.fao.org/aos/agrovoc/c_2187&gt;</li> <li>&lt;http://aims.fao.org/aos/agrovoc/c_2974&gt;</li> <li>&lt;http://aims.fao.org/aos/agrovoc/c_6108&gt;</li> <li>&lt;http://aims.fao.org/aos/agrovoc/c_7152&gt;</li> <li>&lt;http://aims.fao.org/aos/agrovoc/c_7552&gt;</li> <li>&lt;http://aims.fao.org/aos/agrovoc/c_8337&gt;</li> </ul>
prefLabel	<ul style="list-style-type: none"> <li>Jagung (ms)</li> <li>Kukurydza (ziarno) (pl)</li> <li>Mais (it)</li> <li>Mais (de)</li> <li>Maíz (es)</li> <li>Mais (fr)</li> <li>Milho (pt)</li> <li>Misir (tr)</li> <li>kukorica (hu)</li> <li>kukurica siata (sk)</li> <li>kukuřičné zrno (cs)</li> <li>maize (en)</li> <li>кукуруза (зерно) (ru)</li> <li>ذرة صفراء (ar)</li> <li>ذرت (fa)</li> <li>मक्का (hi)</li> <li>ข้าวโพด (th)</li> </ul>

Parte final da página do AGROVOC sobre o milho, destacando o *link* para

exportação dos respectivos códigos dos dados na linguagem RDF/XML.

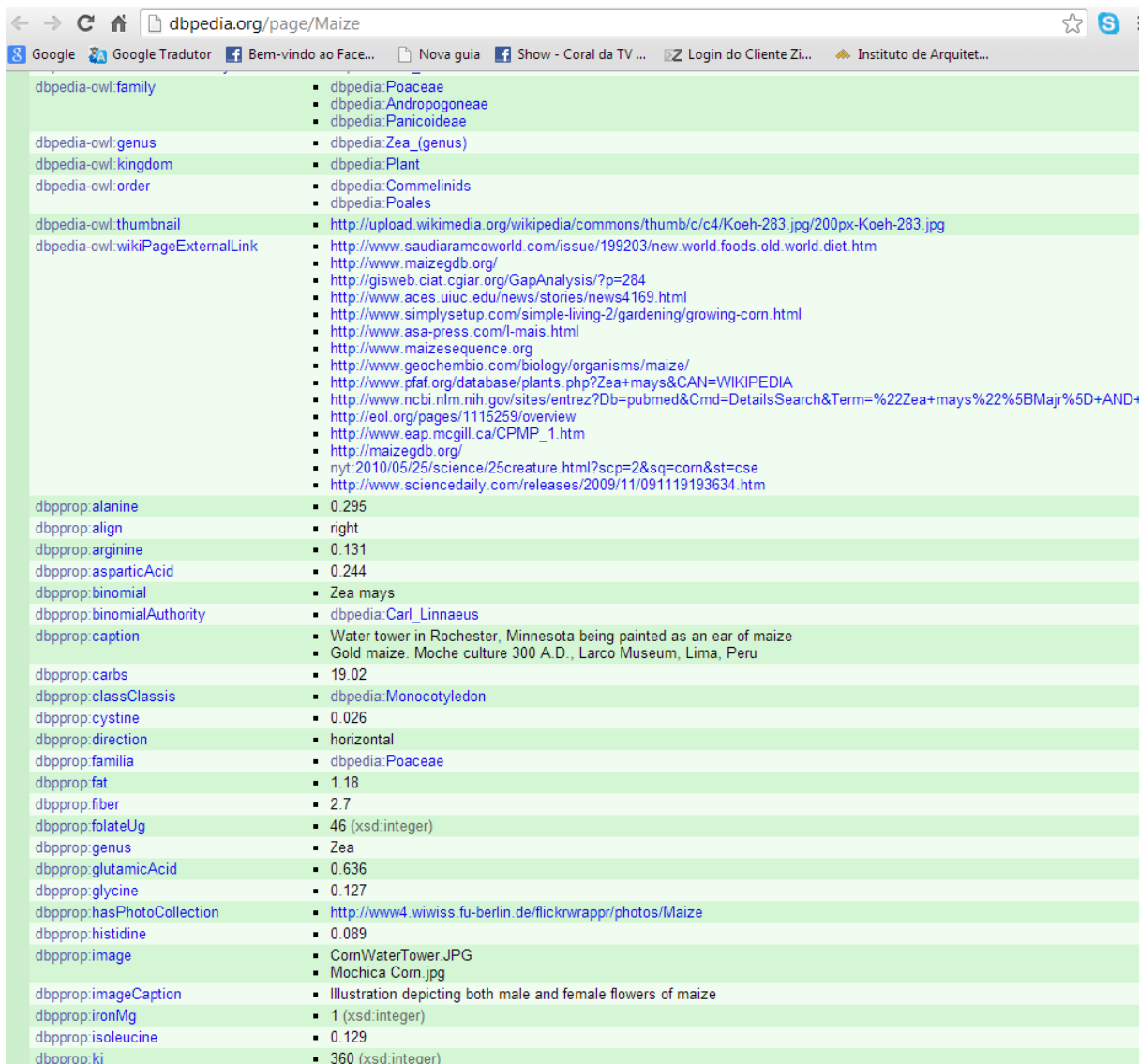
The screenshot shows a web browser window with the URL `aims.fao.org/aos/agrovoc/c_12332`. The page displays a table of properties and values for the term 'Maize'. The 'altLabel' and 'prefLabel' rows are highlighted in green. At the bottom, there is a footer with the text 'The international agricultural thesaurus AGROVOC is maintained by FAO and part of AIMS. The AGROVOC LOD services are a collaboration project with the MIMOS-Rhd.' and two links: 'As Turtle' and 'As RDF/XML', with the latter circled in red.

Property	Value
altLabel	<ul style="list-style-type: none"> <li>Granoturco (it)</li> <li>Ziarno kukurydzy (pl)</li> <li>corn (maize) (en)</li> <li>kukonica (hu)</li> <li>zrno (kukurica) (sk)</li> <li>zrno kukuřice (cs)</li> <li>зерно кукурузы (ru)</li> <li>अनाज (मक्का) (hi)</li> <li>コーン (ja)</li> <li>苞谷 (zh)</li> </ul>
broader	<ul style="list-style-type: none"> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_1474">http://aims.fao.org/aos/agrovoc/c_1474</a>&gt;</li> </ul>
closeMatch	<ul style="list-style-type: none"> <li>&lt;<a href="http://dbpedia.org/resource/Maize">http://dbpedia.org/resource/Maize</a>&gt;</li> </ul>
created	<ul style="list-style-type: none"> <li>2011-11-20T20:35:16Z (dateTime)</li> </ul>
exactMatch	<ul style="list-style-type: none"> <li>&lt;<a href="http://d-nb.info/gnd/4037135-9">http://d-nb.info/gnd/4037135-9</a>&gt;</li> <li>&lt;<a href="http://eurovoc.europa.eu/1744">http://eurovoc.europa.eu/1744</a>&gt;</li> <li>&lt;<a href="http://zbw.eu/stw/descriptor/14093-4">http://zbw.eu/stw/descriptor/14093-4</a>&gt;</li> </ul>
hasStatus	<ul style="list-style-type: none"> <li>Published (string)</li> </ul>
isProducedBy	<ul style="list-style-type: none"> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_8504">http://aims.fao.org/aos/agrovoc/c_8504</a>&gt;</li> </ul>
isUsedToMake	<ul style="list-style-type: none"> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_15500">http://aims.fao.org/aos/agrovoc/c_15500</a>&gt;</li> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_25472">http://aims.fao.org/aos/agrovoc/c_25472</a>&gt;</li> </ul>
modified	<ul style="list-style-type: none"> <li>2012-06-26T22:11:59Z (dateTime)</li> </ul>
narrower	<ul style="list-style-type: none"> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_14385">http://aims.fao.org/aos/agrovoc/c_14385</a>&gt;</li> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_2187">http://aims.fao.org/aos/agrovoc/c_2187</a>&gt;</li> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_2974">http://aims.fao.org/aos/agrovoc/c_2974</a>&gt;</li> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_6108">http://aims.fao.org/aos/agrovoc/c_6108</a>&gt;</li> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_7152">http://aims.fao.org/aos/agrovoc/c_7152</a>&gt;</li> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_7552">http://aims.fao.org/aos/agrovoc/c_7552</a>&gt;</li> <li>&lt;<a href="http://aims.fao.org/aos/agrovoc/c_8337">http://aims.fao.org/aos/agrovoc/c_8337</a>&gt;</li> </ul>
prefLabel	<ul style="list-style-type: none"> <li>Jagung (ms)</li> <li>Kukurydza (ziarno) (pl)</li> <li>Mais (it)</li> <li>Mais (de)</li> <li>Maiz (es)</li> <li>Mais (fr)</li> <li>Milho (pt)</li> <li>Misir (tr)</li> <li>kukonica (hu)</li> <li>kukurica siata (sk)</li> <li>kukuřičné zrno (cs)</li> <li>maize (en)</li> <li>кукуруза (зерно) (ru)</li> <li>ذرة صغيرة (ar)</li> <li>ذرت (fa)</li> <li>मक्का (hi)</li> <li>ข้าวโพด (th)</li> <li>ማጃ (lo)</li> <li>トウモロコシ (ja)</li> <li>玉米 (zh)</li> <li>옥수수 (ko)</li> </ul>

The international agricultural thesaurus AGROVOC is maintained by FAO and part of AIMS. The AGROVOC LOD services are a collaboration project with the MIMOS-Rhd.  
[As Turtle](#) | [As RDF/XML](#)



Página da DBPedia referenciada pela página do AGROVOC referente ao termo “Milho”.



The image shows a screenshot of a web browser displaying the DBpedia page for 'Maize'. The browser's address bar shows 'dbpedia.org/page/Maize'. The page content is a list of properties and their corresponding values, organized into sections. The properties are listed on the left, and their values are listed on the right. The values are often hyperlinks to other DBpedia pages or external resources.

dbpedia-owl:family	<ul style="list-style-type: none"> <li>dbpedia:Poaceae</li> <li>dbpedia:Andropogoneae</li> <li>dbpedia:Panicoidae</li> </ul>
dbpedia-owl:genus	<ul style="list-style-type: none"> <li>dbpedia:Zea_(genus)</li> </ul>
dbpedia-owl:kingdom	<ul style="list-style-type: none"> <li>dbpedia:Plant</li> </ul>
dbpedia-owl:order	<ul style="list-style-type: none"> <li>dbpedia:Commelinids</li> <li>dbpedia:Poales</li> </ul>
dbpedia-owl:thumbnail	<ul style="list-style-type: none"> <li><a href="http://upload.wikimedia.org/wikipedia/commons/thumb/c/c4/Koeh-283.jpg/200px-Koeh-283.jpg">http://upload.wikimedia.org/wikipedia/commons/thumb/c/c4/Koeh-283.jpg/200px-Koeh-283.jpg</a></li> </ul>
dbpedia-owl:wikiPageExternalLink	<ul style="list-style-type: none"> <li><a href="http://www.saudiaramcoworld.com/issue/199203/new.world.foods.old.world.diet.htm">http://www.saudiaramcoworld.com/issue/199203/new.world.foods.old.world.diet.htm</a></li> <li><a href="http://www.maizegdb.org/">http://www.maizegdb.org/</a></li> <li><a href="http://gisweb.ciat.cgiar.org/GapAnalysis/?p=284">http://gisweb.ciat.cgiar.org/GapAnalysis/?p=284</a></li> <li><a href="http://www.aces.uiuc.edu/news/stones/news4169.html">http://www.aces.uiuc.edu/news/stones/news4169.html</a></li> <li><a href="http://www.simplysetup.com/simple-living-2/gardening/growing-corn.html">http://www.simplysetup.com/simple-living-2/gardening/growing-corn.html</a></li> <li><a href="http://www.asa-press.com/l-mais.html">http://www.asa-press.com/l-mais.html</a></li> <li><a href="http://www.maizesequence.org">http://www.maizesequence.org</a></li> <li><a href="http://www.geochembio.com/biology/organisms/maize/">http://www.geochembio.com/biology/organisms/maize/</a></li> <li><a href="http://www.pfaf.org/database/plants.php?Zea+mays&amp;CAN=WIKIPEDIA">http://www.pfaf.org/database/plants.php?Zea+mays&amp;CAN=WIKIPEDIA</a></li> <li><a href="http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubmed&amp;Cmd=DetailsSearch&amp;Term=%22Zea+mays%22%5BMajr%5D+AND+">http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubmed&amp;Cmd=DetailsSearch&amp;Term=%22Zea+mays%22%5BMajr%5D+AND+</a></li> <li><a href="http://eol.org/pages/1115259/overview">http://eol.org/pages/1115259/overview</a></li> <li><a href="http://www.eap.mcgill.ca/CPMP_1.htm">http://www.eap.mcgill.ca/CPMP_1.htm</a></li> <li><a href="http://maizegdb.org/">http://maizegdb.org/</a></li> <li><a href="http://www.nytimes.com/2010/05/25/science/25creature.html?scp=2&amp;sq=com&amp;st=cse">nyt:2010/05/25/science/25creature.html?scp=2&amp;sq=com&amp;st=cse</a></li> <li><a href="http://www.sciencedaily.com/releases/2009/11/091119193634.htm">http://www.sciencedaily.com/releases/2009/11/091119193634.htm</a></li> </ul>
dbpprop:alanine	<ul style="list-style-type: none"> <li>0.295</li> </ul>
dbpprop:align	<ul style="list-style-type: none"> <li>right</li> </ul>
dbpprop:arginine	<ul style="list-style-type: none"> <li>0.131</li> </ul>
dbpprop:asparticAcid	<ul style="list-style-type: none"> <li>0.244</li> </ul>
dbpprop:binomial	<ul style="list-style-type: none"> <li>Zea mays</li> </ul>
dbpprop:binomialAuthority	<ul style="list-style-type: none"> <li>dbpedia:Carl_Linnaeus</li> </ul>
dbpprop:caption	<ul style="list-style-type: none"> <li>Water tower in Rochester, Minnesota being painted as an ear of maize</li> <li>Gold maize. Moche culture 300 A.D., Larco Museum, Lima, Peru</li> </ul>
dbpprop:carbs	<ul style="list-style-type: none"> <li>19.02</li> </ul>
dbpprop:classClassis	<ul style="list-style-type: none"> <li>dbpedia:Monocotyledon</li> </ul>
dbpprop:cystine	<ul style="list-style-type: none"> <li>0.026</li> </ul>
dbpprop:direction	<ul style="list-style-type: none"> <li>horizontal</li> </ul>
dbpprop: familia	<ul style="list-style-type: none"> <li>dbpedia:Poaceae</li> </ul>
dbpprop:fat	<ul style="list-style-type: none"> <li>1.18</li> </ul>
dbpprop: fiber	<ul style="list-style-type: none"> <li>2.7</li> </ul>
dbpprop:folateUg	<ul style="list-style-type: none"> <li>46 (xsd:integer)</li> </ul>
dbpprop:genus	<ul style="list-style-type: none"> <li>Zea</li> </ul>
dbpprop:glutamicAcid	<ul style="list-style-type: none"> <li>0.636</li> </ul>
dbpprop:glycine	<ul style="list-style-type: none"> <li>0.127</li> </ul>
dbpprop:hasPhotoCollection	<ul style="list-style-type: none"> <li><a href="http://www4.wiwiwiss.fu-berlin.de/flickrwrappr/photos/Maize">http://www4.wiwiwiss.fu-berlin.de/flickrwrappr/photos/Maize</a></li> </ul>
dbpprop:histidine	<ul style="list-style-type: none"> <li>0.089</li> </ul>
dbpprop:image	<ul style="list-style-type: none"> <li>ComWaterTower.JPG</li> <li>Mochica Corn.jpg</li> </ul>
dbpprop:imageCaption	<ul style="list-style-type: none"> <li>Illustration depicting both male and female flowers of maize</li> </ul>
dbpprop:ironMg	<ul style="list-style-type: none"> <li>1 (xsd:integer)</li> </ul>
dbpprop:isoleucine	<ul style="list-style-type: none"> <li>0.129</li> </ul>
dbpprop:kj	<ul style="list-style-type: none"> <li>360 (xsd:integer)</li> </ul>