

**Universidade de Brasília - UnB**

Faculdade de Economia, Administração, Contabilidade e Ciência da Informação  
Departamento de Ciência da Informação e Documentação  
Programa de Pós-Graduação em Ciência da Informação

**Autoria de documentos para a Web Semântica: um ambiente de produção de  
conhecimento baseado em ontologias**

**Edgard Costa Oliveira**

**Brasília, 2006.**



Universidade de Brasília - UnB

Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação

Departamento de Ciência da Informação e Documentação

Programa de Pós-Graduação em Ciência da Informação

## **Autoria de documentos para a Web Semântica: um ambiente de produção de conhecimento baseado em ontologias**

**Edgard Costa Oliveira**

Tese desenvolvida no Programa de Pós-Graduação em Ciência da Informação, como requisito parcial para obtenção do grau de doutor.

Orientação: Prof. Dr. Mamede Lima-Marques  
Universidade de Brasília, Departamento de Ciência da Informação e Documentação

Supervisão no exterior: Prof. Dr. Frank van Harmelen  
Vrije Universiteit Amsterdam, Departamento de Ciência da Computação – Inteligência Artificial

**Brasília, 2006.**

© 2006 Edgard Costa Oliveira.

Todos os direitos reservados.

Universidade de Brasília

Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação

Departamento de Ciência da Informação e Documentação

Oliveira, Edgard Costa.

Autoria de documentos para a Web Semântica: um ambiente de produção de conhecimento baseado em ontologias./

Edgard Costa Oliveira. – Brasília : Universidade de Brasília, 2006. (Tese de Doutorado)

260p.

1. Produção de informação. 2. Ontologia. 3. Web Semântica.
4. Autoria de documento.

Tese de Doutorado apresentada ao Departamento de Ciência da Informação e Documentação da Universidade de Brasília, como requisito parcial para a obtenção do grau de Doutor em Ciência da Informação.

Brasília, 04 de agosto de 2006.

Aprovado por:

Prof. Dr. Mamede Lima-Marques – Presidente  
Universidade de Brasília

Prof. Dr. Marcello Bax – Membro  
Universidade Federal de Minas Gerais

Prof. Dr. Murilo Silva de Camargo – Membro  
Universidade de Brasília

Profa. Dra. Marisa Bräscher Basílio de Medeiros – Membro  
Universidade de Brasília

Prof. Dr. Antônio Lisboa Carvalho de Miranda – Membro  
Universidade de Brasília

Prof. Dr. Jaime Robredo – Suplente  
Universidade de Brasília

Dedico este trabalho a meus pais: minha tábua de salvação.

*Quando a palavra emudece e se torna escrita, é aí que fala mais alto e fica ainda mais bonita.*

Eloí Calage

## **Resumo**

Nesta tese, apresentamos o momento da autoria de documentos como situação potencial para a representação do conhecimento e a conseqüente publicação de documentos. Sistemas baseados em ontologia para a representação da informação e autoria na Web Semântica estão surgindo e provendo soluções como anotação semântica, extração da informação, marcação ontológica, etc. Ao analisar esses sistemas, descobrimos as características que eles deveriam ter, a fim de que produzam e suportem documentos em conformidade com as recomendações da Web Semântica e conseqüentemente usufruam de suas potencialidades. Apresentamos aqui uma análise de ferramentas baseadas em ontologias, a partir de um critério de análise específico desenvolvido para identificar as características, requisitos e conformidade com as recomendações existentes, identificadas na revisão de literatura. O resultado da análise revelou que nesse novo paradigma de produção, os autores deverão desempenhar tarefas de estruturação do conhecimento com os quais eles ainda não estão familiarizados. Identificamos a necessidade de abordar essas questões no âmbito da Ciência da Informação, buscando dar apoio a autores no desenvolvimento de conteúdos bem estruturados e representados, a partir do potencial que têm as atuais tecnologias baseadas em ontologias. Após tratar dessas questões, propomos com esse trabalho um conjunto de requisitos e um modelo de arquitetura de ambiente de autoria baseado em ontologia para a Web Semântica, composto por módulos que sugerem novas funções e integram tecnologias correntes, fazendo uso de possibilidades existentes e propondo o desenvolvimento de novos requisitos.

## **Palavras-chave**

Produção de documento, Web Semântica, ontologia, arquitetura da informação, representação do conhecimento, ambiente de autoria.

## **Abstract**

In this thesis we focus on the authoring moment as a potential situation for knowledge representation and the following document publishing task. Ontology-based systems are being developed providing solutions such as semantic annotation, information extraction, ontological markup, etc. By analyzing these systems, developed to the Semantic Web, we intend to find out what the requirements are for these systems to produce and support documents compliant with the Semantic Web recommendations. We present here an analysis of some ontology-based systems, based on specific analysis criteria developed in order to identify features, requirements and compliance with existing recommendations and guidelines identified in the literature review. The results showed that, in this new knowledge production paradigm, authors are asked to perform additional knowledge structuring tasks that they were not familiar with before. We noticed the need to approach these questions in the light of Information Science, especially if we wish to support authors in the development of more structured and represented contents, with the use of ontology technologies. By tackling the question, we propose here a set of requirements and a model of architecture for an ontology-based authoring environment for the Semantic Web, composed of modules that suggest new functions and integrate new and existing technologies, making use the current possibilities and proposing the development of new requirements.

## **Keywords**

Document production, Semantic Web, ontology, information system architecture, knowledge representation, authoring environment.

## Agradecimentos

Ao caro Professor Dr. Mamede Lima-Marques (UnB) pela orientação, paciência, apoio e energia durante todo o trabalho.

À Professora Dra. Marisa Bräscher (UnB) pelas referências e pelo ponta-pé inicial, que me indicou a Web Semântica como ambiente que viabilizaria a idéia deste trabalho.

Ao Professor Dr. Marcello Bax (UFMG) pelas precisas contribuições, apoio e presença na banca examinadora.

Ao Professor Dr. Frank van Harmelen, pela acolhida em Amsterdam, pela confiança na idéia e inúmeras contribuições para o desenvolvimento desta tese.

Aos professores Dr. Antônio Miranda, Dr. Jaime Robredo, Dra. Suzana Mueller, Dr. Emir Suaiden, pela inspiração e liderança durante os meus oito anos de pós-graduação na UnB.

À querida Shirlei de La Cruz Rosas, analista de sistemas que me ajudou na modelagem, com paciência, dedicação e doçura.

Aos colegas de sala de aula do CID que, fizeram dessa jornada acadêmica algo muito maior que uma pesquisa, em especial Alexandra, Cristiane e Herbbertt.

Aos professores, funcionários e alunos do Departamento de Ciência da Informação e Documentação da Universidade de Brasília.

À Professora Dra. Sely Costa por todo o apoio, atenção e sua brilhante presença.

À minha amada mãe, Dona Therezinha e ao meu pai, Seu José Oldaci, pelo amor, carinho, atenção e educação.

Ao meu irmão Cadmo pelo auxílio na tradução e revisão desta tese, originalmente escrita em língua inglesa.

Ao CNPq e à CAPES pelas bolsas de estudos concedidas, em especial ao Programa de Pós-Graduação com Estágio no Exterior-PPGEX.

Ao meu amigo Professor Dr. Kim Veltman pela inspiração, incentivo e ombro amigo durante meu tempo na Holanda, e por oferecer sua vasta biblioteca em Maastricht.

Aos amigos de Amsterdam Duda Paiva, Karen, Pauli (*in memorian*), Fernando, Fré, Paul e Zharko.

A todos os meus amigos que compreenderam minha brevidade durante esses tempos, em especial Paloma, Patrick, Led, Valteno, Marine, Oto, Léo, Laninha, Fabiano, Tânia, toda a Casa da Vó Laila, Ana, Soraya, Blanche, Cruvinel, Bethina, Miriam, Bruno, Simone, Sabrina e todos os demais, tão queridos.

A toda a minha família que entendeu mas muito brigou pela minha presença e energia, em especial Thereza, Carolina, Mariana, Guilherme, Iza, Lúcia, Cláudia, Martin e Jan.

Um agradecimento especial ao vovô Dr. Edgard, *in memorian*.

A Deus, pelo conhecimento e luz amplamente ofertados.

# Sumário

<b>Capítulo 1</b> .....	<b>14</b>
<b>Introdução</b>	
1.1 Objetivos .....	21
1.2 Justificativa da pesquisa .....	22
1.3 Metodologia da Pesquisa .....	25
1.3.1 Ferramentas selecionadas .....	26
1.3.2 Critério de Análise .....	27
1.3.3 Referencial Teórico e Metodológico .....	27
1.3.4 Contribuições da Tese para a Ciência da Informação .....	29
<b>Capítulo 2</b> .....	<b>33</b>
<b>Revisão de literatura</b>	
2.1 Autoria de Documentos .....	34
2.1.1 Introdução .....	34
2.1.2 O modelo cognitivo de escrita .....	35
2.1.3 Escrita e autoria .....	38
2.1.4 Ambiente de autoria .....	42
2.1.5 O Hipertexto .....	45
2.2 Web Semântica .....	47
2.2.1 Documentos compostos .....	51
2.2.2 RDF .....	52
2.2.3 XML .....	54
2.3 Ontologias .....	56
2.3.1 Aplicabilidades das ontologias .....	58
2.3.2 Visualização de ontologias .....	59
2.4 Tecnologias baseadas em ontologia para a Web Semântica .....	64
2.4.1 Anotação Semântica de Documentos .....	65
2.4.2 Geração de Metadados .....	68
2.4.3 Links entre termos de documentos e ontologias externas .....	70
2.4.4 Explicitando a estrutura do documento para revisão de pares .....	73
2.4.5 Conclusões .....	76
2.5 Modelagem de dados em UML .....	79
2.5.1 Diagrama de caso de uso .....	79
2.5.2 Diagrama de Seqüência .....	80
2.5.3 Diagrama de Atividades .....	82
2.5.4 Especificação de Requisitos do Sistema .....	83
<b>Capítulo 3</b> .....	<b>86</b>
<b>Análise de ambientes baseados em ontologias</b>	
3 Análise de Ambientes de Autoria Baseados em Ontologias .....	87
3.1 Critérios de análise dos ambientes de autoria .....	88
3.2 Critérios para análise de tecnologias baseadas em ontologia .....	89
3.3 Critérios para análise de conformidade com recomendações do W3C .....	92
3.4 Critérios para análise de ambientes de software .....	95
3.5 Proposta de critério de análise de ferramentas de autoria baseadas em ontologias .....	97
3.6 Aplicação do modelo nos sistemas analisados .....	99
3.7 Resultados encontrados .....	103

<b>Capítulo 4 .....</b>	<b>107</b>
<b>Proposta de Arquitetura</b>	
4. Uma proposta de arquitetura de ambientes de autoria para a Web Semântica .....	108
4.1 Especificação do nível de meta-modelagem da arquitetura .....	111
4.2 Especificação do Nível de Modelagem .....	113
4.3 Especificação do Nível de Aplicação da Arquitetura .....	117
<b>Capítulo 5 .....</b>	<b>124</b>
<b>Conclusões</b>	
6. Bibliografia .....	129
Anexo I – Compilação dos critérios de análise de requisitos .....	140
Anexo II - Especificação de Requisitos do Sistema DocOnto.....	144
A. Introdução .....	144
A.1 Objetivos da especificação de requisitos .....	144
A.2 Escopo .....	145
A.3 Definições, acrônimos e abreviações .....	145
A.4 Visão geral da especificação de requisitos .....	148
B. Perspectiva do Ambiente .....	148
C. Funções do produto especificado.....	150
D. Características dos usuários do sistema.....	151
E. Limitações do ambiente .....	152
F. Suposições e Dependências .....	152
G. Supressão de requisitos .....	153
H. Requisitos específicos .....	153
H.1 Interfaces externas .....	154
H.2 Funções do software .....	155
I. Requisitos de Desempenho .....	171
J. Conformidade com padrões do W3C .....	171
J.1 Requisitos para documentos em XML .....	172
J.2 Requisitos para ontologias em OWL.....	174
J.3 Requisitos para Documentos compostos.....	176
J.4 Requisitos de Acessibilidade.....	178
J.5 Requisitos para ambiente independente de dispositivo.....	181
K. Atributos do software .....	183
L. Organizando os requisitos específicos .....	184
Anexo III Diagramas do Sistema em UML .....	185
III.1 Modelo de Domínio .....	185
III.2 Diagrama de caso de uso do sistema.....	186
III.3 Diagrama de atividades do sistema .....	188
III.4 Diagramas de seqüência do sistema .....	190

## Lista de Figuras

Figura 1	Elementos do ciclo da informação, 30
Figura 2	Cinco camadas da arquitetura da Web Semântica, 49
Figura 3	Ambiente de sala de aula, 50
Figura 4	Modelo de expressão em RDF, 54
Figura 5	Exemplo de ontologia, 60
Figura 6	Exemplo de taxonomia, 60
Figura 7	Visualização de ontologia, 62
Figura 8	Tela do sistema Protegé, 62
Figura 9	Tela do sistema Magpie, 66
Figura 10	Tela do sistema Gate, 67
Figura 11	Tela do sistema Ellogon, 68
Figura 12	Tela do sistema Cohse, 72
Figura 13	Tela do sistema Claimaker, 74
Figura 14	Tela do sistema Trellis, 75
Figura 15	Protótipo de um modelo de especificação de requisitos, 85
Figura 16	Modelo genérico de arquitetura da informação, 109
Figura 17	Modelo de arquitetura de ambiente de autoria, 111
Figura 18	Modelo geral de arquitetura de ambiente de autoria, 119
Figura 19	Tela do ambiente de autoria para a Web Semântica, 122, 149
Figura 20	Tela do editor do ambiente de autoria, 150

## **Lista de Tabelas**

Tabela 1	Modelo de análise de requisitos, 98
Tabela 2	Consolidação da análise de ferramentas, 102
Tabela 3	Compilação de critérios da análise de requisitos, 141
Tabela 4	Interfaces externas do ambiente de autoria 154

## **Lista de Diagramas do Sistema**

1. Diagrama de modelo de domínio, 185
2. Diagrama de caso de uso do ambiente de autoria, 115, 186
3. Diagrama de atividades do sistema, 116, 188
4. Diagrama de seqüência – Configurar ambiente padrão, 190
5. Diagrama de seqüência – Alterar configuração do ambiente padrão, 191
6. Diagrama de seqüência – Marcar documento, 192
7. Diagrama de seqüência – Procurar as ontologias que contêm o termo, 193
8. Diagrama de seqüência – Associar termo à ontologia, 194
9. Diagrama de seqüência – Desassociar termo da ontologia, 195
10. Diagrama de seqüência – Visualizar ontologias que contêm o termo, 195
11. Diagrama de seqüência – Criar ontologia, 196
12. Diagrama de seqüência – Atualizar ontologia, 197
13. Diagrama de seqüência – Buscar ontologia associada ao termo, 198
14. Diagrama de seqüência – Enviar para validação as alterações de ontologias, 199
15. Diagrama de seqüência – Procurar ontologias via ontobroker, 200
16. Diagrama de seqüência – Definir metadados do documento, 200
17. Diagrama de seqüência – Associar ao termo as fontes de informação, 201
18. Diagrama de seqüência – Navegar semanticamente na ontologia, 202
19. Diagrama de seqüência – Definir assunto do documento, 203
20. Diagrama de seqüência – Incorporar termo da ontologia ao texto, 204
21. Diagrama de seqüência – Validar a ontologia do documento, 205
22. Diagrama de seqüência – Atualizar a ontologia do documento, 206
23. Diagrama de seqüência – Associar fontes à ontologia do documento, 207
24. Diagrama de seqüência – Incluir novas fontes de informação, 207

## **Capítulo 1**

### **Introdução**

## 1. Introdução

As atuais tecnologias de informação e comunicação têm revolucionado as etapas de produção, representação e disseminação do conhecimento. Os editores de textos eletrônicos, por exemplo, permitem uma flexibilidade de registro da informação jamais vista em todos os anos anteriores da história da escrita humana. Com a Internet e suas possibilidades, esses editores foram se adaptando a uma série de novas tecnologias que dão maior suporte aos autores para que possam escrever e estruturar conteúdos, editar textos em vários padrões e plataformas de comunicação e leitura, criação de hipertextos, indexação automática, edição de imagens, entre muitas outras possibilidades.

O advento do hipertexto e a Internet trazem um problema de representação: os *links* entre os documentos são feitos por relações sintáticas. A pesquisa da informação em meio eletrônico está restrita a busca por semelhança de escrita entre os termos. Por exemplo, ao buscar-se manga no Google, obtêm-se manga fruta e manga de camisa.

Com o advento da Web Semântica proposta por Tim BERNERS-LEE (2001), um novo modelo de representação do conhecimento, baseado em ontologias, está buscando formas de interligar os sistemas de informação por meio de indicadores semânticos, construtos artificiais (KLEIN 2004), baseados em inferências e contextos. Graficamente, as ontologias parecem árvores de conceitos, em que cada folha é um termo específico relacionado a um galho que é um conceito mais geral. Serve para mostrar, por exemplo, que manga é uma fruta tropical, e não manga de camisa, num dado contexto.

Na Web Semântica, o conhecimento é explicitado (formalmente) com ajuda das ontologias, que são estruturas semânticas utilizadas para representar o conteúdo de um documento de forma organizada e bem estruturada (BERNERS-LEE, 2001). O uso das ontologias durante o momento da autoria é um assunto ainda em discussão e estudo.

A Web Semântica está sendo estruturada atualmente pelo W3C – World Wide Web Consortium – um consórcio internacional da Web formado por instituições de pesquisa e liderado por Berners-Lee, cujos grupos de trabalho buscam respostas e recomendações sobre como os computadores podem dar suporte à produção e gestão de enormes quantidades de informação na nova Web.

Para que possam estar disponíveis na nova Web Semântica e usufruir de suas possibilidades, sistemas e fontes de informação em linha devem estar de acordo com as recomendações do W3C, mais especificamente na forma com que os conteúdos são representados para dar suporte à interpretação e interoperabilidade automáticas. As ontologias para a Web Semântica têm sido consideradas ferramentas potenciais para o estabelecimento de uma linguagem comum para a descrição e entendimento de conjuntos específicos de conhecimento de uma organização.

Nesse contexto surgem novas formas de gestão da informação, levando-se em conta aspectos cruciais como escalabilidade, validade da informação, integridade das referências, múltiplas fontes de conhecimento, diversidade de conteúdo, *link* e inferência, dentre outros (van HARMELEN, 2002).

Apesar de haver inúmeras pesquisas nas áreas de processamento, disponibilidade, pesquisa, armazenamento e recuperação da informação, há muito para ser feito ainda no tocante a soluções de autoria ou produção de informação. Um novo paradigma de produção reside em novas possibilidades de autoria de documentos, que deixou de ser linear e tornou-se multilinear, baseada em *hiperlinks*. Hoje em dia, os ambientes de edição de textos, sejam .doc, html, xml, java, etc., compartilham de ambientes de rede, permitindo consultas e *links* a documentos e fontes disponíveis em linha, muito embora ainda de forma restrita e dependente do alcance e possibilidade de integração entre as atuais tecnologias.

Sob esse paradigma, os ambientes de autoria de documentos para a Web Semântica estão sendo preparados para ser baseados em ontologias. Assim, a representação semântica dos documentos pode ser feita durante o processo de autoria, e não apenas depois do documento ter sido publicado.

Inúmeros são os benefícios que esta atividade gera, destacando-se a navegação conceitual para recuperação da informação com base na relevância e pertinência dos conceitos contidos em documentos, além dos processos correntes de representação da informação (indexação, marcação, anotação etc.).

Além da questão da interoperabilidade semântica, a questão idiomática é também muito importante, uma vez que aspectos culturais, lingüísticos e de formatação são inerentes às características de cada ambiente de autoria. Os usuários desses ambientes devem ser apoiados a produzir documentos em seus contextos específicos, idiomas e formatos. A Web Semântica está buscando formas

de permitir que os sistemas estejam adaptados às necessidades de cada perfil lingüístico e temático dos usuários no mundo (W3C, 2005).

Um aspecto a ser considerado também é o da produção de conhecimento entre diferentes domínios e também entre diferentes atividades. Cada atividade humana tem suas próprias características e necessidades de comunicação, o que leva a diferentes tipos de informação a ser produzida e utilizada. Por exemplo, a autoria de material técnico-científico (como artigos, trabalhos técnicos, livros) difere em grande escala do material artístico ou comercial (obras, propagandas, filmes). No entanto, nesta tese é apresentado um modelo de ambiente de autoria eletrônico que pode ser utilizado para a elaboração de qualquer tipo de informação, ou seja, um ambiente genérico para qualquer atividade de comunicação escrita social ou individual, seja via um software específico ou coleção de componentes de software. Uma coleção de componentes de software é qualquer produto de software utilizado em conjunto (módulo do sistema, *plug-in*, etc.) ou separadamente (um editor de marcação, editor de imagem, etc.), independentemente de haver colaboração formal ou não entre os desenvolvedores desses produtos (BERNERS-LEE, 2001).

Assim, acreditamos que serão desenvolvidas tantas ferramentas e ontologias em número proporcional aos diferentes tipos de necessidades de autoria, idiomas, perfil de usuários, tarefas, plataformas tecnológicas e linguagens de representação da informação. Com a corrida para o acesso livre e gratuito a softwares e bases de dados públicos, desenvolvimento caseiro de softwares por conta das iniciativas do software livre, muitas empresas, fábricas de software, pesquisadores e desenvolvedores independentes passam a construir suas próprias soluções. Assim sendo, esta tese busca auxiliar esses grupos a estarem conscientes das possibilidades de suas próprias soluções, com base nas melhores práticas das tecnologias atuais. Como fazer com que eles estejam munidos de requisitos específicos que farão com que a produção, disseminação, recuperação e uso da informação sejam mais efetivos e processados pela Web Semântica?

Os grupos de trabalho envolvidos na implantação da Web Semântica, membros do W3C, assim como toda a comunidade de pessoas que fazem a engenharia de ontologias, são incentivados a desenvolver pesquisas que se referem à definição de padrões e critérios para análise e desenvolvimento de ontologias e suas tecnologias correlatas (ANGELE, 2003). No entanto, os critérios existentes até hoje geralmente referem-se apenas à avaliação de ferramentas de edição de

ontologias. Novas ferramentas de autoria baseadas em ontologia estão sendo desenvolvidas sem que haja um conjunto de requisitos ou coleção de componentes que estejam de acordo com a filosofia da Web Semântica.

Por meio desta pesquisa, buscamos preencher essa lacuna e contribuir para a análise de ferramentas baseadas em ontologias, provendo um critério próprio de análise, desenvolvido especificamente para esse propósito. O critério sugerido tem como referência critérios de análise e avaliação relacionados e utilizados para fins correlatos. A pesquisa busca identificar as características comuns entre as ferramentas e projetar novas tendências para ambientes de autoria para a Web Semântica. Os critérios referem-se a quesitos como usabilidade e utilidade dos ambientes de autoria e edição, tecnologias de compartilhamento de conhecimento, gerenciamento de ontologias, ambientes de software assim como documentação de referência. Buscamos assim apontar para um novo modelo de autoria de documentos, considerando o uso simultâneo de tecnologias de ontologias para aprimorar a escolha de termos e conceitos, a criação de metadados e, assim, criar uma estrutura de *hiperlinks* dos conceitos utilizados.

Para compreender como ocorre o processo de autoria à luz da Web Semântica e das ferramentas baseadas em ontologias, buscamos definir e caracterizar o que é autoria baseada em ontologias; definir um critério específico de avaliação de ferramentas existentes; assim como motivar o desenvolvimento de novos ambientes, ao propor uma arquitetura de ambiente de autoria baseado em ontologias.

Também buscamos contribuir para a maturidade do assunto, levantando questões tais como:

- i) quais são as características das ferramentas de autoria existentes e quais são as novas possibilidades de autoria de documentos para a nova Web Semântica?
- ii) quais são os requisitos que esses ambientes devem ter para usufruir das possibilidades que a Web Semântica oferece?
- iii) é possível especificar um modelo de autoria que apresente as características necessárias para uso na Web Semântica?

Considerando-se que a Web Semântica é um novo ambiente para a autoria de documentos, quais seriam os impactos desse novo modelo de representação do conhecimento no momento da autoria? Podem os sistemas e ferramentas de autoria

de hoje em dia ser escaláveis para soluções que são usadas correntemente pelos usuários, quando produzirem documentos sobre qualquer assunto representado por uma ontologia?

Buscamos nesse trabalho responder a essas perguntas e apresentar uma proposta para beneficiar autores de documentos e reduzir o esforço necessário para representar documentos com metadados. De acordo com Waard (2005), é possível que o momento da autoria de um documento seja uma oportunidade única para exportar as novas idéias de um autor para o ambiente eletrônico. A questão é como os ambientes de autoria podem permitir que autores criem ou manipulem entidades e relações entre os conceitos, de modo a representar o que eles querem dizer, por meio de ontologias.

Para apresentarmos os resultados desta pesquisa de doutorado, dividimos o trabalho em capítulos e seções que refletem o desenvolvimento da idéia central da tese, a autoria de documentos para a Web Semântica. O caráter inédito da tese reside na proposta de um ambiente novo jamais especificado, elaborado à luz de técnicas da representação da informação e do conhecimento e de referências científicas para análise e especificação de sistemas.

No Capítulo I está o problema da pesquisa em questão, explicitando os objetivos geral e específicos, propostos para contribuir na solução do problema, assim como a justificativa e metodologia para atingi-los. Nosso objetivo final é propor um modelo de ambientes de autoria que utilizam as ontologias para acessar e representar o conhecimento na Web Semântica no momento em que autores escrevem os documentos.

O Capítulo II apresenta uma revisão de literatura em produção de documentos, onde é feito um levantamento de definições e também a descrição da tarefa à luz de trabalhos de pesquisa atuais na área. Também apresentamos o que é a Web Semântica, ontologias e formatos de documentos, em conjunto com os desafios tecnológicos para colocar em prática essa nova Web baseada em relações semânticas entre os conteúdos, via ontologias. Finalizamos esse capítulo com a apresentação dos fundamentos da metodologia, utilizados para a modelagem e especificação do sistema proposto. Nesse capítulo mostramos o estado da arte em tecnologias baseadas em ontologia para Web Semântica, assim como algumas ferramentas de autoria. As tecnologias oferecem opções de anotação semântica de documentos, de geração de metadados, de *links* entre termos de documentos e entre si, e finalmente apresentamos tecnologias que explicitam a estrutura retórica de documentos.

No Capítulo III, é apresentado o primeiro resultado da pesquisa, dividido em duas seções: critério de análise de ferramentas de autoria baseadas em ontologias; e análise de ambientes com base no critério proposto. Analisamos seis ferramentas de autoria e apresentamos os resultados obtidos.

No Capítulo IV é apresentado o resultado final da pesquisa, o objetivo geral atingido por meio da proposta de especificação de alto nível de arquitetura de ambientes de autoria da Web Semântica. A arquitetura está representada por meio de uma especificação de requisitos e modelagem do sistema, detalhado no anexo II e anexo III desta tese.

Finalizamos o trabalho no Capítulo V, com a conclusão alcançada, de que é possível juntar em um novo ambiente de produção de documentos os ambientes de autoria, de ontologias e de navegação semântica. O trabalho se amplia com especificação e modelagem do sistema, constantes nos anexos, assim como a bibliografia utilizada para a realização da pesquisa.

## **1.1 Objetivos**

Os ambientes de autoria de documentos baseados em ontologia indicam um caminho promissor para a produção do conhecimento, pois visam facilitar e automatizar processos de produção, representação, disseminação e recuperação de informação, durante a autoria do documento, conforme apresentamos no desenvolvimento desta tese. Diante desse contexto, a presente pesquisa possui três objetivos específicos que compõem o objetivo geral a seguir.

### **1.1.1 Objetivo Geral**

Criar uma especificação de alto nível de um novo modelo de arquitetura de sistemas de informação para ambientes de autoria de documentos baseados em ontologia para a Web Semântica.

### **1.1.2 Objetivos Específicos**

Para alcançar o objetivo geral desta tese, os seguintes objetivos específicos foram buscados:

- i) Levantar critérios e modelos de análise de sistemas de autoria, de avaliação de ontologias, de conformidade com o W3C e de análise e avaliação de software.
- ii) Realizar uma análise de requisitos de sistemas de autoria baseados em ontologia com base em critérios levantados.
- iii) Especificar requisitos e modelar a arquitetura de um novo ambiente de autoria baseado em ontologias para a Web Semântica.

## **1.2 Justificativa da pesquisa**

Alguns grupos estão trabalhando em formas de representação da informação no momento da autoria do documento (OLIVEIRA, 2004). No entanto, ainda não estão organizados e sistematizados os requisitos necessários para que esses ambientes se adequem às necessidades dos usuários, aos quesitos de funcionalidade e que estejam de acordo com as recomendações do W3C.

Para tanto, são necessárias novas abordagens modulares no design de software que possam fornecer ferramentas de autoria que sejam apropriadas, flexíveis e adequadas às diferentes preferências e práticas dos autores de documentos eletrônicos. No entanto, acredita-se que o desenvolvimento de ferramentas adequadas só será possível se os escritores profissionais e os pesquisadores de formas de escrita compreender e participar dos processos de desenvolvimento de softwares (SHARPLES, 1996).

Com o desenvolvimento progressivo da Web Semântica e a necessidade de representação do conhecimento de acordo com as recomendações do W3C, muitas organizações e grupos de pesquisa passam a desenvolver suas próprias ferramentas. No entanto esses grupos precisam ser guiados durante essa tarefa, de modo a contar com o auxílio necessário para construir esses ambientes de acordo com as recomendações propostas na literatura referente e pelo W3C. Uma vez que novos ambientes de autoria estão surgindo no mundo, critérios de avaliação devem estar disponíveis para auxiliar na avaliação de funcionalidades e na composição dessas novas ferramentas.

Até então, os ambientes de autoria e os ambientes de pesquisa eletrônica são dois mundos distintos. O que se propõe nesta tese é que esses dois ambientes podem ser integrados, gerando informações que possam ser processadas de forma adequada pela Web Semântica. Para tanto, propomos o uso das ontologias como mecanismos de orientação e aquisição do conhecimento por parte dos autores para que gerem representações semânticas da informação, durante a produção do documento, ou seja, durante a autoria.

A melhoria que o uso de ontologias traz ao acesso e representação do conhecimento durante o momento da autoria é uma questão que está sendo estudada por alguns grupos de pesquisa, conforme apresentado nesta tese. Isso se reflete no cenário atual do desenvolvimento do que chamamos de ambientes de

autoria baseados em ontologias. Para atingir esses objetivos, os grupos estão desenvolvendo ferramentas que usam ontologias para oferecer novas possibilidades de escrita, demonstrando que os autores passam a ser os principais responsáveis pela representação de seus próprios documentos.

Ocorre que os autores ainda não estão preparados para realizar tarefas diferentes das que realizam hoje para gerar e publicar um documento na Web. Logo, esta tese busca apresentar um modelo de arquitetura de um ambiente de autoria que ofereça facilidade de uso e familiaridade com funções já desempenhadas correntemente, adicionando-se apenas o trabalho de consulta a ontologias durante a autoria, e realização de *links* entre os termos às fontes de informação, via ontologias.

Segundo KIM (2002), as ontologias serão amplamente utilizadas caso sejam desenvolvidas por qualquer produtor de conhecimento, não apenas por ontologistas. Para ele, as ferramentas de ontologias precisam de fato ser avaliadas de acordo com critérios de facilidade de uso e capacidade de expressar conceitos ricos mesmo sem a necessidade de conhecimento aprofundado em representação do conhecimento por parte dos autores, independentemente de serem ou não usadas para partilhar dados. KIM (2002) afirma que as ontologias serão também usadas para especificação de softwares.

No entanto, não se sabe ainda quais são os requisitos que devem ser atingidos por essas ferramentas de modo que elas usufruam dos potenciais da Web Semântica. Consideramos que devem ser desenvolvidos ambientes de produção que gerem documentos que estejam de acordo com as recomendações do W3C para representação da informação. Ademais, há uma ausência de critérios de avaliação de ferramentas de autoria, considerando as inúmeras possibilidades de produção que elas podem oferecer aos autores.

É sabido que o atual paradigma de publicação eletrônica vem mudando drasticamente a forma com que as pessoas produzem conhecimento e conseqüentemente na forma com que elas disponibilizam esse conhecimento por meio dos documentos. Isso está refletido no aparecimento de um conjunto de funcionalidades para a autoria de documentos que ainda não está muito claro para todos. Buscamos, assim, identificar e apresentar essas funcionalidades, de forma a propor requisitos que permitam um aproveitamento maior das possibilidades que a Web Semântica oferece. Assim, os desenvolvedores de novos ambientes de autoria

para a Web Semântica poderão prover tecnologias e serviços voltados aos autores, com base em critérios sistematizados e em conformidade com o estado da arte, assim como as recomendações do W3C.

Iniciativas como a do WAIG - Grupo de Trabalho em Acessibilidade do W3C (W3C, 2005) têm sido aplicadas com sucesso no desenvolvimento de orientações para desenvolvedores de ferramentas de autoria no que tange à acessibilidade por parte de usuários com necessidades especiais. No entanto, não há até então no W3C, ou em qualquer outro organismo de regulamentação, abordagens que revelem requisitos gerais ou específicos que levem em consideração o uso, interoperabilidade, etc., de forma a dar suporte ao desenvolvimento de ferramentas de autoria de documentos que façam uso do potencial de representação do conhecimento que a Web Semântica oferece.

A contextualização de parâmetros de qualidade é uma questão chave para as novas gerações de consumidores e produtores de mídias eletrônicas (VELTMAN, 1997). A presente tese busca essa contextualização, uma vez que encontramos uma enorme carência, na área, de orientações sobre parâmetros de qualidade e requisitos de funcionalidade para o desenvolvimento de novas e atuais ferramentas.

A avaliação de ontologias e suas tecnologias é um aspecto crítico para a integração destas soluções em aplicações reais, pois caso as definições da ontologia e os ambientes em que elas são usadas não forem suficientemente avaliados, a comunicação entre os agentes dos sistemas pode não ocorrer adequadamente (GÓMEZ-PÉREZ, 1995).

### **1.3 Metodologia da Pesquisa**

Para o alcance do objetivo geral e de cada um dos objetivos específicos propostos nesta pesquisa, apresentamos nesta seção as metodologias respectivamente adotadas. A proposta metodológica geral desta pesquisa foi publicada na 3ª Conferência Internacional da Web Semântica ISWC/2004, em Hiroshima, Japão, intitulada: “*A framework for ontology-based authoring environments*” (OLIVEIRA, van HARMELEN e LIMA-MARQUES, 2004).

**Objetivo específico i) Levantar e sistematizar critérios e modelos de análise e avaliação de sistemas.**

**Metodologia:** Pesquisa exploratória de critérios de análise e avaliação relacionados ao tema de pesquisa, divididos nos seguintes grupos: a) critérios de análise de sistemas, b) critérios para análise e avaliação de tecnologias de autoria e c) critérios de análise de tecnologias baseadas em ontologia; d) critérios de conformidade com as recomendações do W3C.

Para fins de análise de requisitos das ferramentas identificadas, desenvolvemos um critério específico para a análise de ambientes de autoria baseados em ontologia, uma vez que não existe até o momento esse tipo de critério ou metodologias que abordem o problema. Para tanto, foram estudados alguns critérios e abordagens existentes, os quais foram adaptados para a análise ora realizada. Essas metodologias foram agrupadas e associadas a teorias e literaturas referenciadas, e foram selecionados os aspectos necessários de acordo com as possibilidades e objetivos de um novo ambiente de autoria.

**Objetivo específico ii) Realizar uma análise de requisitos de sistemas de autoria baseados em ontologia com base em critérios levantados.**

**Metodologia:** A análise foi conduzida a partir da sistematização de ferramentas de autoria identificadas na revisão de literatura. Para análise, foi utilizado o critério resultante do objetivo específico 1, o qual foi utilizado para sistematizar características presentes ou não nas ferramentas analisadas. O resultado deste trabalho foi publicado em OLIVEIRA (2006).

### **Objetivo específico iii) Especificar requisitos e modelar uma arquitetura de um novo ambiente de autoria baseado em ontologias**

**Metodologia:** Especificação de alto nível de um modelo de arquitetura para ambientes de autoria da Web Semântica – a arquitetura proposta foi desenvolvida com base em uma metodologia apresentada a seguir:

a) Especificação de requisitos com base na norma *IEEE Std 830-1998 / IEEE Recommended Practice for Software Requirements Specifications*, elaborada pelo *Software Engineering Standards Committee do Institute of Electrical and Electronics Engineers* (IEEE, 1998) *Computer Society* e utilizada amplamente como referência para o desenvolvimento de sistemas.

b) Elaboração de Diagramas em UML a partir dos requisitos levantados:

- Modelo de Domínio;
- Diagrama de Caso de Uso;
- Diagrama de Atividades do Sistema;
- Diagramas de Seqüência de Atividades do Sistema.

A proposta metodológica desta tese foi aceita pela Vrije Universiteit de Amsterdam, como tema de projeto de pesquisa de mestrado em Ciência da Computação, e está disponível no endereço [www.few.vu.nl/krr](http://www.few.vu.nl/krr), intitulado “*Research Proposal of an Ontology-based Authoring Environment*” (OLIVEIRA, 2004). Essa será uma das alternativas para a viabilização do desenvolvimento do ambiente de autoria para a Web Semântica proposto nesta tese, e sua conseqüente implementação em trabalhos futuros.

#### **1.3.1 Ferramentas selecionadas**

Os critérios para seleção das ferramentas, de acordo com a definição de ambientes de autoria para a Web Semântica, apresentada no item 1.3.3, baseiam-se nas atividades que essas ferramentas desempenham: autoria, anotação semântica e *hiperlink*. As seguintes ferramentas foram identificadas na revisão de literatura e aqui apresentadas:

- S-Cream Ont-o-Mat (HANDSCHUH, 2002);
- WICKED (CARR 2004);
- Trellis (GIL, 2002);
- Claimaker (UREN, 2003); e
- Semantic Word (TALLIS, 2003).

### **1.3.2 Critério de Análise**

Para realizar a análise das ferramentas, definimos um modelo próprio de análise a partir de um conjunto de critérios metodológicos para identificação de funcionalidades e requisitos, agrupados em quatro macro-categorias:

- características do ambiente de autoria;
- características do gerenciamento da ontologia;
- características de conformidade com as recomendações do W3C;
- características de engenharia de software.

A ausência de abordagens para a análise de ambientes de autoria para a Web Semântica levou-nos a desenvolver um critério específico de análise, baseado em abordagens similares existentes em áreas relacionadas (ex: engenharia de software), e com base no estado da arte e nas recomendações do W3C, consultadas e adaptadas de MACKNIGHT (1989), GRUBER (1993), DAVIS (1993), GÓMEZ-PÉRES (1995), ZELKOWITZ (1998), USCHOLD (1999), TALLIS (1999), ONTOWEB (2002 e 2004), HANDSCHUH (2002), DENNY (2002 e 2004), ISO (2001), ANGELE (2001 e 2003), CRISTEA (2004), OLIVEIRA (2004) e W3C (1999-2006).

### **1.3.3 Referencial Teórico e Metodológico**

A construção do referencial teórico e metodológico desta tese de doutorado iniciou-se com a pesquisa e o estudo de literatura disponível sobre os assuntos relacionados à autoria, nas áreas da Ciência da Informação, da Ciência da Computação e Lingüística, que permitiram a organização da informação sobre a autoria de documentos em ambiente eletrônico. Esse estudo, apresentado na revisão de literatura (Capítulos 2 e 3), forneceu elementos para a construção de um referencial teórico e metodológico. O conteúdo da revisão de literatura foi redigido a partir dos seguintes tópicos.

- A perspectiva cognitiva do processo de escrita apresentado primeiramente por FLOWER e HAYES (1977) e incrementado por GEEST e SHARPLES (1996), a qual deu base para o que se conhece hoje como ambientes eletrônicos de autoria.

- A descrição de processos históricos de produção de informação, textos e de documentos, que deram a base para o paradigma atual dos editores de textos e

hipertextos, a partir das visões de Vannevar Bush e seu *Memex* (1945), e dos conceitos de hipertexto de Douglas Englebart e Ted Nelson.

- O paradigma atual da representação do conhecimento na Web Semântica, suas características e linguagens de representação, como o RDF e XML, que formam a base para a construção e uso das ontologias.

Durante o estágio de doutorado realizado na Vrije Universiteit de Amsterdam, tivemos contato com membros do W3C, pesquisadores da representação do conhecimento e desenvolvedores de ambientes baseados em ontologia para a Web Semântica. O resultado dessa experiência permitiu a sistematização de práticas existentes e o conhecimento dos requisitos de sistemas que buscam apresentar soluções ao problema da autoria na Web Semântica, resultando nos tópicos apresentados no Capítulo 3 desta tese:

- as tecnologias baseadas em ontologia para a Web Semântica, onde identificamos e analisamos sistemas com propósitos comuns ao da autoria de documentos na Web Semântica, publicado em OLIVEIRA (2004).

- a sistematização de características dos sistemas e a definição e explicitação do novo conceito de “*ambientes de autoria baseados em ontologia*”, publicado em OLIVEIRA (2004).

Esse trabalho permitiu a identificação de um referencial teórico que guiou a realização desta pesquisa, de acordo com a prevalência de enfoques teóricos e empíricos estudados. Desta forma, chegamos à seguinte definição de um ambiente de autoria baseado em ontologia para a Web Semântica (OLIVEIRA, 2004):

*“um conjunto de ferramentas de escrita, edição e representação de documentos que interativamente apóiam os usuários (autores), permitindo um melhor acesso, uso e representação semântica do conhecimento durante a autoria, por meio das seguintes tarefas:*

- *fazer anotação semântica de documentos;*
- *criar metadados;*
- *ligar os termos do documento com as ontologias externas;*
- *ligar documentos similares entre si;*
- *transformar citações em links rotulados;*
- *tornar explícita a estrutura retórica de um documento; e*
- *ligar documentos a comentários de revisão de pares.”*

O referencial metodológico para análise de sistemas baseados em ontologia resultou da análise de metodologias e critérios correlatos, divididos em três grandes categorias:

- critérios de análise de sistemas de informação;
- critérios de análise de tecnologias de autoria;
- critérios de análise de tecnologias baseadas em ontologias;
- critérios de análise de tecnologias da Web Semântica.

O resultado dessa análise de critérios proveu a base para a proposta e uso de um conjunto de requisitos específicos para a análise de características de ambientes de autoria baseados em ontologia para a Web Semântica, objetivo específico alcançado por esta pesquisa, apresentado no Capítulo 3 e publicado em OLIVEIRA (2006).

O referencial metodológico para o desenho do ambiente de arquitetura teve como referência inicial o modelo genérico de arquitetura da informação proposto por LIMA-MARQUES (2005), descrito no capítulo 3, e publicado em OLIVEIRA (2006).

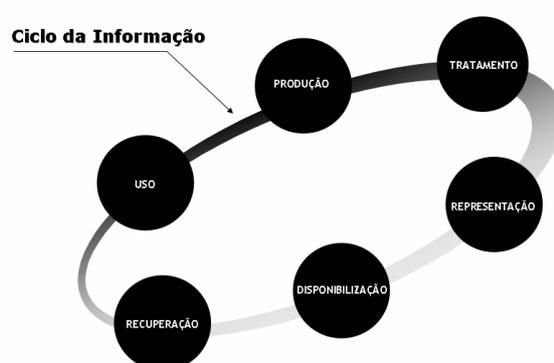
O referencial metodológico para a especificação de requisitos e modelagem do sistema teve como referência e formato a norma IEEE830 - *System Requirement Specification* (1998), descritos no anexo II. A especificação de requisitos de conformidade com a Web Semântica teve como base o estado da arte em autoria de documentos, em recomendações e notas do W3C, os quais foram agregados à proposta de requisitos formulada nesta tese. A linguagem utilizada para a modelagem do sistema foi a UML – *Uniform Modelling Language* (2006), referência para a construção dos desenhos dos diagramas de representação do ambiente de autoria, apresentados no anexo III.

#### **1.3.4 Contribuições da Tese para a Ciência da Informação**

A humanidade encontra-se atualmente em um momento de revolução científica (KHUN, 1983) que afeta diretamente as formas de se fazer pesquisa científica, caracterizada por novas formas de disseminação do conhecimento. Esse é um momento de reavaliação das formas de se fazer ciência, onde novas formas de produção e apresentação da informação estão emergindo (KIRCZ, 1998), o chamado modo II de produção científica de Gibbons (1994). É também o que podemos chamar de *ciberiência* (NENTWICH, 2003), que é identificada pelo

desenvolvimento de novas formas de pesquisa. Nesse cenário, outras ferramentas estarão disponíveis para autores produzirem uma quantidade cada vez maior de conhecimento e tornar disponível a literatura científica, sob o impacto das novas tecnologias da informação e comunicação, o que nos remete às previsões de Vannevar Bush (1945) e de Martin Greenberger com os “*computadores do amanhã*” (1964).

Essas questões são objeto de estudo da Ciência da Informação, a qual tem como universo as seis principais etapas do ciclo da informação (LE COADIC, 1996):



**Figura 1 – Elementos do ciclo da informação.**

As etapas de tratamento, representação, disponibilização, recuperação e uso do ciclo da informação têm sido amplamente abordadas nos âmbitos da Ciência da Informação e da Ciência da Computação, como revela o universo de pesquisa corrente que se pode ter acesso. No entanto, há uma enorme carência de estudos específicos, teóricos e práticos, sobre a etapa de produção da informação.

Atualmente, os ambientes de produção, pesquisa e representação da informação atuam de maneira independente e pouco integrada. A proposta de ambiente de autoria apresentada nesta tese busca aliar autoria com pesquisa e representação da informação, ou seja, oferecer ao usuário (autor) mecanismos para enriquecer a representação da informação e a pesquisa de fontes de informação durante o processo de produção de um documento, com a ajuda de ontologias.

Estudos futuros poderão contribuir para corroborar a hipótese de que a representação da informação, quando realizada pelo autor durante a autoria,

umenta as chances da eficiência do processo de indexação e disponibilização da informação em ambientes eletrônicos.

Assim, a principal contribuição desta tese está em apresentar soluções para facilitar a representação da informação por meio de metadados, descritores de tipos de documentos e sobretudo por meio das ontologias, durante o processo de produção de informação, via documentos eletrônicos. Conseqüentemente, contribui-se para as demais etapas do ciclo: disponibilização, recuperação, uso, produção e tratamento da informação. Para tanto, buscamos nesta tese propor uma solução prática e sistêmica que englobe os assuntos de produção de documentos (autoria), pesquisa e representação da informação, a partir de abordagens tecnológicas oriundas da Ciência da Informação, da Ciência da Computação e da Linguística. Logo, a tese busca apresentar uma abordagem prática voltada a beneficiar os usuários de sistemas de informação, em particular os autores de documentos eletrônicos.

Outra grande contribuição desta tese é trazer para o âmbito da Ciência da Informação no Brasil discussões sobre a viabilização da Web Semântica e inovações para a produção, representação e disponibilização da informação nesse novo ambiente. Esses problemas têm sido abordados intensamente no âmbito da Ciência da Computação, e dada a característica informacional dessa questão, torna-se eminente a sua abordagem à luz da Ciência da Informação. Para SOUZA (2004), todo embasamento filosófico, metodológico e conceitual da Web Semântica parte do núcleo duro da Ciência da Informação. Segundo o autor, tudo indica que os padrões que estão sendo desenhados para esta nova Web também sejam adotados na arquitetura de bibliotecas digitais e de novos sistemas de informação, destacando-se:

- construção de novas interfaces com o usuário para sistemas de informação;
- construção automática de tesouros e vocabulários controlados;
- projetos de novos e melhorados motores de busca;
- indexação automática de documentos;
- gestão do conhecimento organizacional; e
- gestão da informação estratégica e da inteligência competitiva.

Acreditamos que novas frentes de pesquisa se abrirão, em decorrência dos problemas que surgem com o desenvolvimento de novas tecnologias de informação, e que afetarão diretamente as atividades do profissional da informação.

Destarte, buscamos colocar a pesquisa brasileira em Ciência da Informação em posição de vanguarda na Gestão do Conhecimento para a Web Semântica, visto que ela representa hoje um ambiente potencial e também revolucionador das etapas de produção e representação da informação na Web atual.

**Capítulo 2**  
**Revisão de Literatura**

## 2.1 Autoria de Documentos

### 2.1.1 Introdução

O surgimento da escrita foi o marco do desenvolvimento da humanidade. A forma com que essa atividade é realizada hoje (eletrônica) é tão distinta da forma original (pedra cunhada) que a única coisa comum entre elas é a necessidade de registro da informação. Desde a sua invenção, um espectro variado de práticas de escrita foi utilizado pelo homem de acordo com as possibilidades de cada época, baseado nas tecnologias disponíveis.

Tudo começou com a invenção da escrita pelos sumérios nos vales dos rios Tigre e Eufrates, hoje o Iraque. A evolução do livro partiu da tábua de terra cota, dos papiros, do *codex* e chegou ao livro eletrônico (KILGOUR, 1998). Nos idos de 1970, um processador de texto custava em média 15 mil dólares, e nos anos 90 eles estavam disponíveis por 400 dólares e foram vendidas mais de um milhão de cópias. Os processadores de texto têm sido vendidos como sendo um sistema de publicação completo, pois apresentam muitas alternativas de produção de documentos, como por exemplo, redação, edição e formatação de texto, corretor gramatical e ortográfico, tesouro, tipos de fontes e recursos de impressão variados. Os processadores de texto agilizaram a produção de documentos do autor para o editor e do editor para a gráfica, fazendo com que o autor tenha mais responsabilidade sobre seu documento, pois além de redigi-lo, o autor edita e organiza o documento para publicação.

Com o crescimento da pesquisa em hipertexto, multi e hipermídia, o livro eletrônico está em pleno processo de desenvolvimento (KILGOUR, 1998). *“Não é possível antecipar o fim do livro impresso, mas pode-se antecipar a aceitação do livro eletrônico. Não somos capazes de resgatar a mudança que houve entre o uso dos papiros para o codex, mas sabemos que o último conviveu durante muito tempo com o primeiro”*.

Hoje não só mais se dispõem de tábuas, papiros, livros e notebooks. Uma malha de comunicação em rede, interligada por meio de códigos identificadores, permite uma série de novas formas de produção de conhecimento em ambientes eletrônicos de autoria.

### **2.1.2 O modelo cognitivo de escrita**

As existentes abordagens sobre o uso cognitivo da linguagem remontam a Piaget (1982), que considerava a linguagem como uma estrutura *de* e *para* conhecimento. Piaget dizia que o desenvolvimento da linguagem, por ser uma atividade da mente, é indispensável para a elaboração do pensamento. Tida por Vigotsky como um instrumento, a linguagem é crucial para o desenvolvimento cognitivo pois é elaborada socialmente. A escrita torna-se um caminho de entrada e saída para os domínios do saber e aprender (BARRETT, 1998), no entanto, a aplicabilidade de modelos de processos cognitivos na linguagem da escrita eletrônica é ainda um problema em estudo.

Flower e Hayes(1977) foram os primeiros a abordarem o processo de escrita sob a perspectiva cognitiva. Para eles *“escrever requer muito mais do que apresentar palavras em uma página, e muito do tempo do escritor será gasto em gerar idéias, formar planos e intenções, coleccionar recursos e tratar de problemas sobre retórica e estrutura”*. Essa visão influenciou o trabalho de muitos pesquisadores do processo de escrita profissional, conhecido como modelo cognitivo de escrita.

SMITH (1989) também propôs um sistema de escrita baseado no modelo cognitivo de Flower e Hayes, composto por modos cognitivos: o de rede – permite que o escritor explore idéias ao unir frase que indicam conceitos por meio de uma rede gráfica; o modo árvore – usado para organizar a estrutura do documento em forma de hierarquia; o modo editor – que provê um editor de textos padrão; e o modo de texto – para navegação e edição de blocos de textos.

SHARPLES (1992) propôs um sistema similar, que oferece três visões do documento em produção: visão de redes de nós que permitem ao escritor colocar as idéias e ligá-las em uma rede de associações; visão de estrutura que dá ao escritor a habilidade de criar e manipular a estrutura do texto; e a visão linear que permite ao escritor digitar o texto com o mínimo de interferência.

Escrita é um processo que consiste em sub-processos específicos como planejamento, tradução do planejamento em linguagem e revisão da linguagem gerada, e, se necessária, a sua revisão. Esses sub-processos são alternados e interligados, fazendo da tarefa de escrita um processo cognitivo recursivo e iterativo. (GEEST, 1996). Para isso é que foram criados os processadores de texto: ajudam

os escritores a escrever e revisar textos. As facilidades que eles oferecem foram feitas para dar suporte a sub-processos separados: apoio ao planejamento do documento, tesouros para escolha de palavras variadas, corretores de gramática e ortografia que auxiliam na edição dos textos.

O mais importante em um modelo cognitivo é a capacidade de interconexão e alternância entre os sub-processos da redação de textos. Os resultados desses estudos não são freqüentemente aplicados em critérios de design de ferramentas de autoria que foram surgindo amplamente no mercado de softwares. Apesar de existirem tantas bases de conhecimento ao dispor dos pesquisadores, a abordagem cognitiva da escrita não contribuiu muito para o desenho dessas ferramentas (GEEST, 1996). Cabe aos profissionais da informação e aos pesquisadores das tecnologias de produção de documentos, revelar pesquisas que possam dar apoio ao desenvolvimento de avançadas formas de produção com base em modelos cognitivos.

GEEST (1996) realizou uma pesquisa com escritores e atores do processo de produção de documentos, que revelou que o planejamento documentário não é um processo individual, estritamente cognitivo, mas uma etapa na produção de documento que envolve muita comunicação entre os atores envolvidos no processo de escrita, como por exemplo: o cliente para quem o documento se destina; os colegas trabalhando no mesmo documento; o gerente responsável pela qualidade do produto final; e o processo no qual o documento é produzido. Como resultado, o planejamento do documento engloba mais do que o processo cognitivo de definir “o que dizer” e “o que fazer” no texto a ser escrito. Esse estudo também demonstrou que o planejamento e a revisão do documento são duas atividades geralmente separadas no tempo e realizadas por atores diferentes, mas que são intrinsecamente relacionadas.

Com essa pesquisa, Geest propôs um modelo com quatro fases principais no processo de produção de documento, que não são necessariamente lineares, e as atividades inerentes podem ser recursivas. São elas:

- 1a fase: definição e planejamento (estrutura do projeto);
- 2a. fase: produção do texto (definição de estrutura, rascunhos, formatação);
- 3a. fase: avaliação do produto (revisão, prova, pré-teste e aprovação) ;
- 4a. fase: distribuição (publicação, divulgação e manutenção).

SHARPLES (1992) realizou uma pesquisa com uma metodologia sócio-cognitiva para analisar o design de tecnologias de escrita em ambientes de trabalho, com o objetivo de identificar orientações práticas de design para sistemas de autoria destinados aos engenheiros de software. A pesquisa levou em consideração as necessidades dos escritores, questionários de práticas, resultados de experimentos controlados, modelos do processo de escrita e uma descrição do contexto organizacional e social onde se dá a escrita dos documentos. O resultado foi o de que *“novas abordagens modulares no design de software poderiam fornecer ferramentas de autoria que sejam apropriadas, flexíveis e adequadas às diferentes preferências e práticas dos escritores”*. SHARPLES (1992), no entanto, acredita que isso apenas será possível se os escritores profissionais e os pesquisadores de formas de escrita compreenderem e participarem dos processos de desenvolvimento de softwares.

SHARPLES (1992) predisse na época que o escritor do futuro precisará ser hábil não apenas em redação e paginação de textos, mas também em desenho gráfico e edição de vídeo. O elemento principal a ser considerado por designers de novos ambientes de escrita é como fazer melhor uso das tecnologias de forma a melhorar e não dificultar a escrita profissional. Para ele, a escrita poderá tornar-se insuportavelmente complexa e a autoria, em todas as suas formas diversas, pode tornar-se uma profissão extremamente técnica e especializada.

Essa perspectiva histórica do design de softwares justifica a característica das ferramentas comerciais ou públicas hoje disponíveis: de serem modulares; adaptáveis às necessidades dos usuários; e fáceis de serem utilizadas. KIRCZ (1998) sugere o uso de módulos de informação cognitiva para a transferência de informações como uma forma estruturada de publicação eletrônica e possibilidades futuras para se lidar com o excesso de informações.

HARMSZE (1999) propõe uma estrutura modular para a autoria de artigos em ciências experimentais. WALKER (1989) também aborda a modularidade de documentos como formas de se gerenciar a informação de documentos em níveis micro e macro. Devido ao fato de haver similaridades entre a autoria de textos e engenharia e desenvolvimento de sistemas, o autor defende a necessidade de se adotarem procedimentos de engenharia para a construção de documentos: *“os ambientes de autoria de documentos e sistemas podem prover características similares para engenheiros de softwares”*. Com base no conceito de modularidade

(resultado do design de um programa em partes pequenas) e abstração (ato de nomear algo e usá-lo com base no que faz e não no como funciona), o autor aplicou o método em seu sistema, um ambiente de engenharia de software (WALKER, 1989), onde encontrou uma forma de rotular os módulos de documentos, como seções, de forma a representar a informação com profundidade e transformar estes rótulos em *links*.

### **2.1.3 Escrita e autoria**

SHARPLES (1996) define o que é ambiente de escrita, como “*tudo aquilo que está em volta do escritor, incluindo os recursos e ferramentas de redação e os sistemas eletrônicos de criação de documentos*”. A distinção entre a escrita e autoria está cada vez mais tênue, uma vez que as tecnologias atuais oferecem mais possibilidades de apoio ao processo de redação de documentos.

Na época, HASELKORN (1988) apresentou quais seriam as novas tendências dos processos de escrita para o futuro, dentre elas:

- a tomada de consciência por parte dos executivos de que a indústria de computadores não deve ser voltada apenas a questões de engenharia, mas também às necessidades dos usuários;
- o aumento da insatisfação com o processo de criação isolada de documentação, etapa final do desenvolvimento de produtos;
- o aumento do uso de sistemas computacionais como um mecanismo primário para produção de documentos;
- o aumento da consciência e insatisfação com o alto custo da produção de documentos;
- o aumento de métodos formais para relacionar o design de sistemas com questões de perspectiva de usuários como, por exemplo, interface e documentação (usabilidade);
- o aumento da aplicabilidade de técnicas da Inteligência Artificial para dar suporte ao desenvolvimento de sistemas e aos usuários.

Para HASELKORN (1998), os escritores profissionais irão contribuir muito no desenvolvimento de sistemas especialistas ao passo que se beneficiam da disponibilidade desses novos meios de produção de documentos. Algumas mudanças previstas por ele com relação ao desempenho dos escritores foram:

- escritores serão envolvidos cada vez mais cedo nos processos de negócios;
- escritores testarão usabilidade, não apenas para ajustar os documentos, mas para alimentar o design e desenvolvimento com os resultados desses ajustes;
- escritores se tornarão cada vez mais sofisticados tecnicamente (sem sacrificar seu conhecimento especializado em perspectivas de usuários e técnicas de comunicação);
- os escritores desempenharão um papel mais importante no design e desenvolvimento de interfaces de usuários;
- os escritores contribuirão no desenvolvimento de métodos formais para fazer *links* entre a funcionalidade de um programa como a interface e documentação correspondentes. Naturalmente que os escritores vão querer adquirir novas técnicas de trabalho (ex: lingüística computacional) para ajudá-los nessas novas tarefas, mas é importante considerar que essas novas técnicas serão inúteis sem o conhecimento desses escritores no tocante a perspectivas de usuários e situações de comunicação;
- escritores desenvolverão seus trabalhos em processamento de linguagem natural e engenharia do conhecimento para auxiliar no desenvolvimento de sistemas de documentação inteligentes.

Dadas essas perspectivas, Haselkorn conclui que o papel dos escritores na indústria do computador passará por mudanças fundamentais, embora suas funções básicas de comunicação permanecerão as mesmas e que possivelmente no futuro não mais iremos chamá-los de escritores.

Para BOLTER (1991), a escrita é como se fosse um jogo criativo de signos, e os computadores oferecem novas formas de jogar com estes signos, por meio de uma plataforma para gravar e apresentar textos junto a novas tecnologias de organização da informação. Para Bolter, um texto escrito é uma estrutura no espaço que implica uma estrutura no tempo: escrever torna o tempo em espaço.

Para BARRETT (1988), o ato de escrever pode ser considerado não apenas com uma função de suporte de informação, mas principalmente como um contexto mais amplo em que se dá o desenvolvimento das idéias. A vanguarda tecnológica e as tradições da retórica não estão necessariamente em oposição, para tanto, novas definições de meios e finalidades de comunicação são necessários para que se

adquiram novas perspectivas sobre como se dá a pesquisa e escrita de documentos no meio eletrônico.

Para MONACO (1999), autoria é o processo de construção de um CD-ROM ou DVD. Autoria envolve programação, teste e preparação dos arquivos que serão gravados em disco, e não inclui a escrita ou redação da informação. Para o autor, o termo autoria não deveria substituir escrita.

HANSEN (1998) diz que autoria é a criação de um programa interativo de computador com o uso de uma linguagem ou sistema de autoria. Desenvolvedores sem conhecimentos formais de programação podem preparar aplicações para sistemas de computador ou CD-ROM por meio de comandos de *script*. Autoria requer uma abordagem disciplinada no preparo de elementos de um programa de multimídia, por meio de um design e planejamento cuidadosos. Para Hansen, *autor* é a pessoa que participa na criação de um programa multimídia ou CD interativo.

Um sistema de autoria, ainda segundo HANSEN (1998), é um programa que facilita o desenvolvimento de multimídia interativa. Sistemas geralmente variam quanto a capacidade e funcionalidade, e fatores como plataforma, audiência e resultados esperados são considerados quando da seleção de uma ferramenta apropriada. Leva muito menos tempo desenvolver uma aplicação multimídia por meio de um sistema de autoria do que apenas por meio da programação e compilação de códigos. Os métodos usados para desenvolver gráficos, texto, vídeo, áudio, animação e outras mídias geralmente não são afetados pela escolha de um ambiente de autoria. Sistemas designados para a criação e edição de tipos específicos de mídia são usados para refinar objetos que serão importados pelo ambiente de autoria, muito embora uma ferramenta de autoria tenha características rudimentares de edição. Hansen nomeia os seguintes elementos como tipos primários de características de ambientes de autoria: linguagem de *script*, controle de fluxo por ícones, sistema baseado em grafos, objetos hierárquicos, *link* entre hipermídias. Um exemplo desses sistemas são os conhecidos *Authorware* e *Director* da *Macromedia*; *Apple Media Tool and HyperCard* da *Apple*; *Oracle Media Objects*; *Asymetrix' Toolbook*; *Hyperwriter* de *Ntergard*; *CourseBuilder* de *Discovery Systems*, dentre outros (HANSEN, 1998).

DORNER (1992) considera três efeitos do novo processo de autoria com apoio das tecnologias de informação e documentação:

- os ambientes físicos e sociais do escritor estão na esfera eletrônica;
- os computadores estão permitindo novas abordagens de trabalho, com o crescimento da autoria colaborativa distribuída de documentos. A união do processador de texto, do correio eletrônico e do acesso remoto a arquivos permite que os escritores de diferentes locais possam trocar notas e rascunhos de trabalhos em andamento;

- a tarefa de escrita está em transformação por causa das possibilidades e restrições impostas pelos computadores. A dinâmica da escrita deixou de ser uma seqüência de etapas sucessivas e tornou-se um processo contínuo de produção e revisão, com a reutilização de textos prontos, estruturação do documento e escrita multilinear, dentre outros.

DORNER (1992) recomenda aos designers de sistemas, que desejam desenvolver ambientes de escrita mais amigáveis, entenderem que a escrita não é apenas um processo cognitivo, mas também cultural e pertence ao contexto profissional do escritor.

Pesquisas de BEREITER e SCARDAMALIA (1987), revelam que o processo de escrita pode ser mais bem descrito pela forma como o conhecimento do escritor é processado: *“em muitas situações de escrita profissional, os escritores não estão buscando em sua memória de longo prazo as idéias sobre o que dizer, mas sim utilizando-se de fontes de informação externas como resultado de um intenso pensamento reflexivo durante a escrita”*. Eles consideram que o planejamento é um importante momento da cadeia de produção de documentos, onde as linhas do projeto a ser redigido são estabelecidas e os escritores formam as idéias do resultado esperado. Escrita *“não é necessariamente uma atividade individual, ela requer uma interação intensa e recorrente entre clientes, gerentes e co-autores, no questionamento, design, discussão e negociação tanto do produto quanto do processo em que ele é desenvolvido”*.

#### **2.1.4 Ambiente de autoria**

Os editores de texto modernos provêm diversas possibilidades de redação e edição de documentos como o *Microsoft Word* (MICROSOFT, 2003), a plataforma LaTeX (LAMPORT, 1994), dentre outros. Segundo NENTWICH (2003), é fundamental que haja um bom editor de textos, de modo que a edição não distraia o autor, no caso os cientistas, da sua tarefa principal: escrever um texto acadêmico. No futuro, eles poderão auxiliar na organização do hipertexto, provendo assistência ao autor durante a criação de *links* e módulos dos documentos. Além disso, essas ferramentas farão uma identificação por rótulos e armazenamento adequado do documento no ambiente de rede, fazendo *links* semânticos semi-automaticamente (NENTWICH, 2003).

A pesquisa de SHARPLES (1996) resultou na seguinte lista de diretrizes para ambientes genéricos de autoria. Ele também utilizou como referência metodológica os trabalhos de JONES (1995) – diretrizes para o desenvolvimento de ferramentas de escrita colaborativa – e o trabalho de Sharples e Pemberton (1990) – diretrizes para o design de processadores de documentos centrados em usuários:

- operações básicas de edição, incluindo a edição do texto, da rede de notas e da estrutura do documento;
- múltiplas visões do documento em construção: espacial, não linear, linear e formatada;
- fácil movimentação entre as visões;
- criação de marcações informais e rascunhos; *link* entre o texto do documento a um meta-texto (como anotações, por exemplo);
- configuração e manipulação de restrições;
- sobreposição de tarefas e incorporação de outras como criação de notas e envio de e-mail durante o processo de escrita;
- reutilização de produtos anteriores, incluindo o texto, estrutura do documento e de notas;
- comunicação entre autores colaboradores por uma rede de debate e revisão de texto;
- controle de versão;
- mudanças de intenção, incluindo uma grande facilidade de “desfazer” ações;

- união de múltiplas contribuições de construção e revisão de texto;
- acesso ao histórico de interações entre os escritores e o texto;
- alternância de autores participantes do processo.

Ele estabeleceu indicadores de como as diretrizes acima podem ser aplicadas na prática por meio de novas abordagens de design de ambientes de escrita.

1. Transferência – refere-se à facilidade de transferência e compartilhamento de formatos de informação.
2. Sistemas modulares – um software integrado por cada conjunto de funções.
3. Sistemas mínimos – interfaces simples que adotam protocolos familiares aos usuários.
4. Design sócio-cognitivo de sistemas – metodologia de desenvolvimento de tecnologias complexas que refletem o contexto social e cognitivo do ambiente em que se dá a comunicação.

O trabalho de BROCKMANN (1990) apresenta uma lista de ferramentas disponíveis para facilitar a escrita de documentos:

- ferramenta de entrada de dados com modelos, tabelas e padrões de estilo;
- scanner de texto;
- corretor ortográfico;
- tesouro;
- checagem de gramática e pontuação;
- marcadores de texto;
- ferramentas para geração de tabelas e índices de conteúdos;
- ferramentas de autoria de hipertexto;
- ferramentas para inclusão de gráficos;
- tradutores de sgml;
- hardware de alta qualidade.

Considerou três possibilidades de desenvolvimento de processos de autoria de documentos (BROCKMAN, 1990):

- acelerar a velocidade do processo de documentação;
- dar suporte a abordagens centradas no usuário;

- redefinir a noção do que é escrita, o que poderá fazer com que o termo “documentador” (*technical writer*) desapareça.

SHARPLES (1996) afirma que novas formas de design devem contar com resultados de pesquisas e análises sobre as práticas de escrita. “*O desenvolvimento das ferramentas de autoria não é apenas uma questão técnica, mas uma questão de engenharia social e organizacional.*”

SHNEIDERMAN (1989) diz que “*usuários experientes de processadores de texto podem aprender a usar uma ferramenta de autoria em uma hora. No entanto, a sintaxe e a semântica da ferramenta de autoria são apenas a base para a construção de aplicações de hipertexto, assim como aprender a digitar é apenas a base para se escrever um romance ou artigos de periódicos.*” (p.119).

Duas questões importantes no processo de autoria são a gerência dos nós e a indicação de *links* nos hipertextos, pois pode ser muito útil constatar que um sistema de edição de hipertexto possui um índice de nós referenciados ou criados. Além disso, mecanismos automáticos de criação de *links* e marcação de texto podem ser muito atraentes, mas podem distrair o leitor.

SHNEIDERMAN (1989) propõe características básicas de ferramentas de autoria:

- a variedade de funções de edição disponíveis (cópia, inserção, remoção, mudança geral no documento etc.);
- a disponibilidade de uma lista de nomes de *links*, termos indexados, sinônimos etc.;
- a variedade de comandos de edição visual do documento;
- a disponibilidade de funções de busca e substituição para fazer mudanças globais em vários nós;
- controle de cores (texto e fundo);
- capacidade de mudança rápida entre os modos autor e navegador para testar idéias;
- acesso a CD-Rom, vídeo e outros recursos;
- capacidade de exportar arquivos de outros sistemas;
- operabilidade em uma rede local;
- bases de dados distribuídas para multi-usuários em rede;
- controle de versão;
- editores de gráfico e mecanismos para uso de vídeos;
- colaboração;
- compressão de dados;
- controle de segurança;
- criptografia;

- confiabilidade;
- integração com outros softwares e hardwares;
- disposição do navegador.

### **2.1.5 O Hipertexto**

A invenção do hipertexto, viabilizada pela estrutura em rede da Internet, revolucionou as técnicas e possibilidades de escrita e autoria. Vemos na literatura estudada, que a transição do uso do termo autoria em substituição ao termo escrita deu-se com as abordagens teóricas e pragmáticas de hipertexto. Poderíamos de antemão afirmar que este padrão exige mais do que apenas escrever textos, mas de arquiteturas de termos conectados por meio de relações semânticas apresentadas sob diversas formas de visualização da informação. Esse novo cenário de produção é chamado pelos teóricos do hipertexto de ambiente de autoria.

Vários especialistas em hipertexto, como Jacques Derrida e Theodor Nelson, Roland Barthes e Andries van Dam (apud LANDOW, 1997), dentre outros, são unânimes em dizer que devemos abandonar sistemas conceituais fundados nos princípios de centralização, margem, hierarquia e linearidade, e substituí-los pelos princípios da multilinearidade, nós, *links* e redes (LANDOW, 1997). LANDOW lembra que Roland Barthes descreve uma textualidade ideal que é similar ao que propõe o hipertexto: um texto composto de blocos de palavras ou imagens ligadas eletronicamente por múltiplos caminhos, cadeias ou trilhas de uma textualidade aberta, perpetuamente inacabada, descrita pelos termos *link*, nó, rede, Web e caminho. Nesse texto ideal, Barthes (apud LANDOW, 1997) diz que

*“as redes são muitas e interagem sem que elas se sobreponham ao resto; esse texto é uma galáxia de significantes e não uma estrutura de significados, não tem fim, é reversível, e temos acesso a ela de diversas formas, sem que nenhuma delas possa ser estabelecida como a principal; onde os códigos de referência vão tão longe quanto o alcance dos olhos, são indetermináveis (...), a linguagem é infinita”.*

Assim como Barthes, Foucault define o texto em termos de redes e *links*. Na arqueologia do conhecimento, ele indica que as *“fronteiras de um livro nunca são*

*visíveis, pois pertencem a um sistema de referências a outros livros, outros textos, outras frases, é um nó de uma rede, uma rede de referências.*" (FOUCAULT, 1972).

Desde os escritos de Vannevar Bush que é dada muita ênfase ao fato de que as estruturas de hipertextos são construídas de forma similar à cognição humana, ou seja, a organização da memória é vista como uma rede semântica na qual os conceitos são ligados por meio de associações. Se essa premissa estiver válida, SHNEIDERMAN (1989) aborda o hipertexto como sendo uma forma eficiente de aprendizado. Teorias de aprendizado indicam que um hipertexto facilita a aprendizagem pois está mais focado nas relações entre as idéias do que em fatos isolados. As associações que os *links* fazem na base de dados de hipertextos deveriam facilitar a lembrança, a formação de conceitos e o entendimento. (SHNEIDERMAN, 1989). Ademais, um senso de controle que os hipertextos impõem à leitura pode gerar um envolvimento maior com o assunto e uma vontade de ler mais, por parte do usuário.

LANDOW (1997) define hipertexto para denotar um texto composto por blocos de textos (*lexia* para Barthes) e os *links* eletrônicos que os unem. O conceito de hipermídia simplesmente estende a noção de texto em hipertexto, incluindo informação visual, som, animação e outras formas de dados. LANDOW (1997) não faz distinção entre hipertexto e hipermídia, uma vez que, segundo ele, o hipertexto faz *links* de um trecho de discurso verbal para imagens, mapas, diagramas e som tão facilmente como para outros trechos de discursos, expandindo a noção de texto para além de simplesmente verbal.

## 2.2 Web Semântica

O desenvolvimento de tecnologias para a Web tem permitido um maior acesso e disponibilização de informações, como por exemplo os motores de busca como o Google, Alta Vista, Yahoo, dentre muitos outros. No entanto, são conhecidas as limitações desses mecanismos quanto à eficiência e eficácia da recuperação da informação. Vários são os motivos: resultados de busca extensos mas de pouca relevância; informações indisponíveis sobre diversos assuntos; respostas dependentes de precisão do uso de vocabulários específicos (variações lingüísticas e ambigüidades não são tratadas); diversas terminologias não integradas que levam a recursos de informação equivocados; pulverização da informação, distribuída em diversos suportes, levando a uma inevitável extração manual da informação. Ademais, os mecanismos de busca são aplicações isoladas de outras ferramentas, o que torna a difícil integrar recuperação e interpretação da informação (ANTONIOU e van HARMELEN, 2004).

Hoje o líder responsável pela Web Semântica é Tim Berners-Lee, atual presidente do W3C, organização que busca desenvolver tecnologias e padrões, e portanto implementar a Web Semântica. A Web Semântica é uma infra-estrutura visionária, na qual o conhecimento será criado e conectado de uma forma semântica revolucionária (BERNERS-LEE, 1999). Ela está projetada para permitir o uso de sistemas de informação avançados, onde o conhecimento estará organizado em espaços conceituais, com ferramentas automatizadas para checagem de consistência da informação; novas formas de extração de conhecimento; resposta de *queries* de conjuntos de documentos ao invés da pesquisa por palavras-chave de um único documento; e finalmente acesso à informação de acordo com o perfil do usuário.

Os trabalhos de Berners-Lee, também inventor da Web, na época à frente do CERN na Suíça (*Centre des Études Recherches Nucleaires*), estão hoje institucionalizados pelo W3C (2005), uma organização sem fins lucrativos composta por universidades, centros de pesquisa e empresas que apóiam a idéia da Web Semântica.

A Web Semântica representa um universo eletrônico promissor de criação e *hiperlink* entre fontes de conhecimento de forma revolucionária, o que nos remete às

propostas de Vannevar Bush, Douglas Englebart e Ted Nelson. Essa grande estrutura representa uma mudança radical nas formas produção de documentos para a Web, com o crescimento das tecnologias de ontologia que permitiram a anotação da informação e representação, navegação e visualização de informações semânticas, dentre outras. As ontologias surgem como uma forma de associação intuitiva entre os conteúdos da Web.

*“É importante que a Web ajude as pessoas as serem intuitivas e analíticas, porque nossa sociedade necessita dessas duas funções. Os seres humanos têm um equilíbrio natural ao utilizar as partes criativas e analíticas de seus cérebros.”* (BERNERS-LEE, 1999, p. 201.)

Conclui-se que devemos considerar esses assuntos à luz da Ciência da Informação, pois a Web Semântica representa um ambiente rico em questões que lhe são pertinentes, em especial a produção e representação da informação.

A autoria de documentos é apresentada nesta tese como elemento principal para auxiliar na representação da informação por meio de uma infra-estrutura semântica que permite ao autor navegar por fontes de informação, via ontologias, para adquirir conhecimento e fazer *hiperlinks* com o seu documento, durante a autoria.

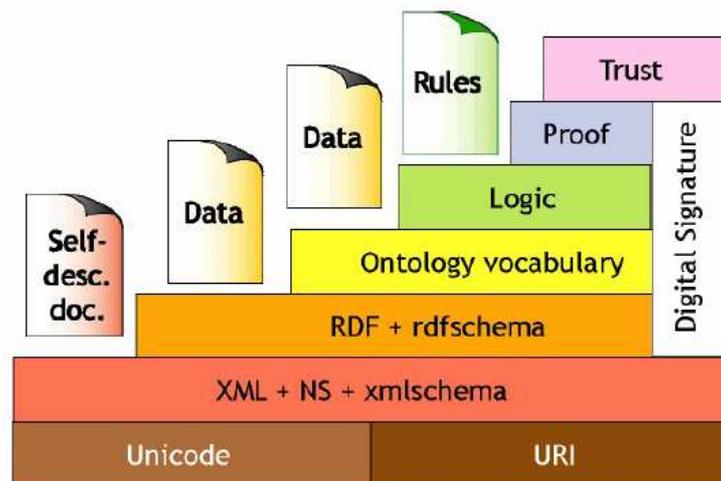
O raciocínio de agentes inteligentes na Web Semântica será possível por meio do uso de duas tecnologias: metadados para identificar e extrair informação dos recursos da Web; ontologias para dar apoio à pesquisa na Web, interpretar informações recuperadas e permitir a comunicação entre outros agentes; e da lógica matemática, para processar a informação recuperada e gerar conclusões (ANTONIOU e van HARMELEN, 2004). Esse raciocínio também pode ser estimulado por meio do uso de tesouros, que podem ajudar na montagem das ontologias e expandir a sua capacidade de representação semântica, utilizando-se também de recursos de representação como o RDF e XML, em total conformidade com a Web Semântica (W3C, 2006).

O XML permite a criação de rótulos (*tags*) dos termos, relacionando-os a categorias que podem ser processadas automaticamente, enquanto que o HTML não oferece essa abordagem. A combinação de XML com RDF e ontologias representa uma alternativa eficiente para essa tarefa.

Cada objeto na Web é identificado por uma URI – *Uniform Resource Identifier*, que são endereços numéricos unívocos. A conhecida URL dos

navegadores é um tipo de URI para páginas Web, sejam elas de texto, vídeo, imagem, som, etc., seguida do tipo de protocolo de comunicação que se utilizada, seja HTTP (páginas da Web), FTP (transferência de arquivos), dentre outros.

Logo, o desenvolvimento da Web Semântica está calcado nos seguintes elementos, que podem ser deduzidos do “*layer cake*” abaixo (bolo de camadas), consagrado para explicitar a proposta da W3C em camadas.



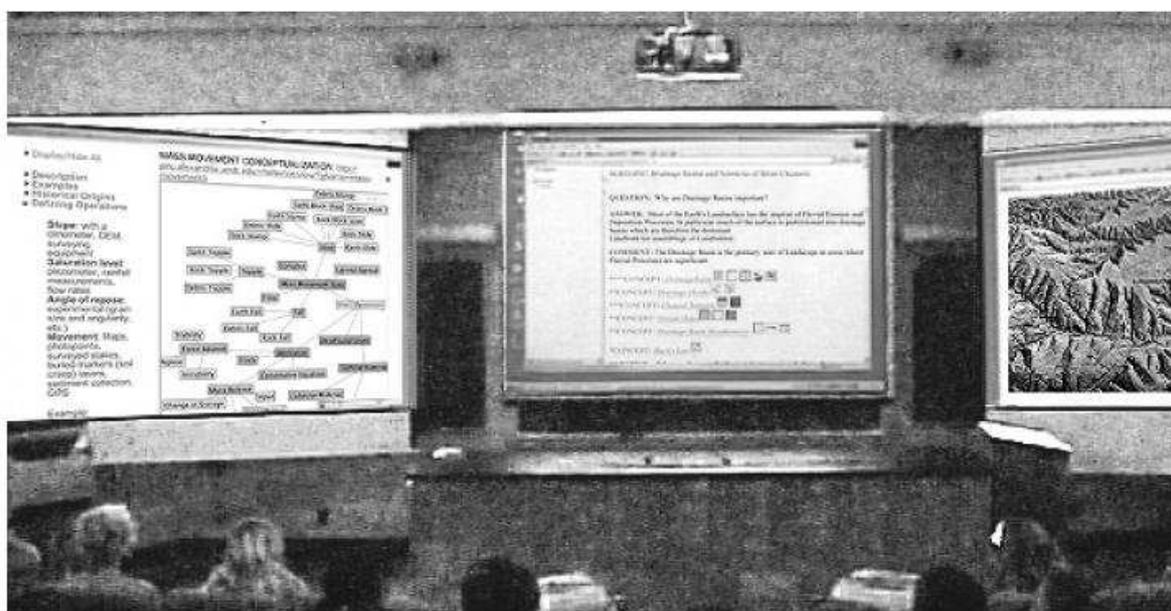
**Figura 2: Camadas da Arquitetura da Web Semântica (BERNERS-LEE, 2001)**

A partir da figura temos:

- **XML** (*Extensive Markup Language*) é uma linguagem de construção e intercâmbio de documentos com um vocabulário definido pelo usuário (autor);
- **RDF** (*Resource Description Framework*) é um modelo de dados para se escreverem afirmações sobre os objetos (recursos).
- **Esquema RDF** fornece os elementos de modelagem: classes e propriedades, relações subclasses e subpropriedades, domínio ou área e restrições.
- **Camada de Lógica**, opera sobre a camada de ontologias, para se construírem relações mais complexas entre os objetos da Web que não são facilmente construídas com o esquema RDF, usada para escrever conhecimentos declarativos específicos à aplicabilidade que se propõe.
- **A Camada de Confiança** (*Trust*) é para ser usada com assinaturas digitais por meio de agentes confiáveis ou por meio de autoridades certificadoras e grupos de usuários da informação, com enfoque em segurança e qualidade da informação provida (ANTONIOU e van HARMELEN, 2004).

SOUZA e ALVARENGA (2004) apontam que os padrões atualmente investigados para dar suporte à Web Semântica também poderão ser adotados para a construção de arquiteturas de bibliotecas digitais e de sistemas de informação. Espera-se, segundo os autores, que o papel do bibliotecário e dos profissionais da informação será diretamente afetado e que mudanças serão necessárias nessas atividades.

Encontramos na revisão de literatura um exemplo ilustrativo de uma idéia de descrição desse novo mundo semântico, projetado para a área educacional, no trabalho de ZENG (2003), a seguir, de um ambiente educacional construído com base em ontologias.



Janela de conceitos

Janela da aula

Janela do objeto

**Figura 3: *Learning Environment Display* – Ambiente de Sala de Aula (ZENG, 2003)**

Concluimos que a proposta da Web Semântica vem revolucionar o processo de representação do conhecimento, fazendo com que as informações hoje disponíveis não mais sejam baseadas na sintaxe, e sim na semântica da informação, expandindo mais ainda a possibilidade de acesso, visualização e uso da informação.

As áreas de comércio eletrônico serão também, naturalmente, beneficiadas por esse ambiente semântico, o que se dá pelo fato de que a Web Semântica se propõe a permitir uma extração da informação mais sofisticada, por meio de agentes

inteligentes que apresentam a informação de uma forma correta e baseada em necessidades específicas de cada usuário e contexto (ANTONIOU e van HARMELEN, 2004).

O objetivo desse novo paradigma é o de viabilizar ambientes de autoria que reflitam uma atitude voltada para o social, que também é atitude da Web Semântica, onde pessoas estarão mais interessadas em trabalhar em conjunto para se comunicarem com o mundo exterior em níveis diferentes. Os documentos produzidos por tais sistemas, que permitem a representação da estrutura semântica de idéias do autor, representam o próximo passo na construção de uma Web comunicativa e compartilhada (BERNERS-LEE, 1999).

### **2.2.1 Documentos compostos**

A Web representa um ambiente heterogêneo que pode comportar a produção de documentos em diversos formatos, como é o caso dos documentos compostos. Um documento composto é um documento que agrega diversas linguagens e recursos por referência (externa) ou por inclusão (contida no documento), como áudio, vídeo, imagem, etc. O W3C (2005) recomenda padrões para a autoria de documentos compostos, os quais podem ser utilizados por ferramentas de autoria e sistemas de produção de conteúdo que dão base às seguintes tarefas:

- edição manual de texto via editor eletrônico;
- geração dinâmica por meio de servidores Web ou outros sistemas *back end*;
- autoria com enfoque em linguagens componentes (XHTML);
- autoria com enfoque em documentos compostos (ferramenta de autoria para CDF<sup>1</sup>).

Espera-se que, com o advento da Web Semântica, devem surgir diversas categorias de ferramentas de autoria para documentos compostos, tais como:

- ferramentas de autoria focadas em documentos para criação de documentos que reúnem textos estáticos e gráficos com o propósito de publicação na Web;
- ferramentas de autoria focadas em multimídia para a criação de conteúdo interativo e *time-based* (como filmes ou jogos, por exemplo);

---

<sup>1</sup> Compound Document Format – Formato de Documentos Compostos

- ferramentas de autoria focadas em aplicação, para a criação de interfaces de usuários;
- ferramentas de autoria focadas em formulários para a criação de modelos de dados;
- ferramentas de autoria independente de dispositivo, para a criação de conteúdo que pode ser adaptado a diferentes requisitos de usuários por meio de diferentes tipos de mídias.

Hoje a Web está sofrendo uma mudança de linguagem: o HTML deixa de ser a linguagem de autoria mais usada, por ser simples demais para a troca de informação, ou seja, não oferece uma estrutura nem uma semântica da informação. O XML passa a ocupar seu lugar, por dar suporte à representação da informação com base em um vocabulário compartilhado e com base no significado da informação, ou seja, nos metadados.

Na Web Semântica, recomenda-se o uso dos padrões de metadados do Dublin Core (DUBLIN CORE, 2004), e representar informações via metadados é um elemento crucial de modo a recuperar informações depois de publicadas (WAARD, 2003). Metadados não são apenas usados para representar descritores bibliográficos, mas também como informação descritiva de um contexto, qualidade e características de um dado (SENSO, 2003). A pergunta que se faz hoje é como os metadados podem ser representados e estruturados pelo próprio autor de um documento ou por engenheiros de conhecimento antes do documento ser publicado?

Antes de nos aprofundarmos nesta questão de ambientes de autoria para Web Semântica, iremos tratar da aplicação das ontologias na representação da informação, via XML e RDF.

### **2.2.2 RDF**

O RDF é uma estrutura de descrição de recurso desenvolvida para representar recursos, modelar metadados e permitir o gerenciamento da informação na Web Semântica (W3C, 2004). O RDF possui uma sintaxe resumida que reflete um modelo de dados simples baseado em gráficos e semântica formal (baseada na lógica matemática) com noções de ligação entre os elementos definidas rigorosamente, com o objetivo de prover uma base para deduções bem estruturadas

de dados em RDF. O desenvolvimento do RDF é motivado pelos seguintes casos de uso:

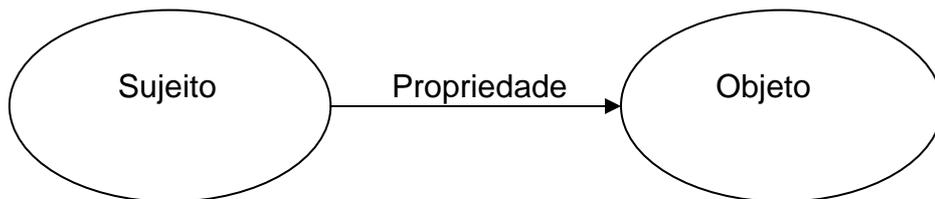
- metadados para a Web – fornecer informações sobre recursos da Web e seus respectivos sistemas (descrição de funcionalidades, conteúdos, preferências, etc.);
- aplicações que requerem modelos de informação abertos, em vez de limitados (fechados), como processos organizacionais, anotação de páginas Web, etc.;
- tratamento de informações processadas automaticamente na Web do mesmo modo com que se faz com o hipertexto, permitindo que os dados processados fora de um ambiente específico no qual foram criados possam ser utilizados em toda a Internet;
- interconexão entre diversas aplicações da Web, combinando dados dessas aplicações e formando novas informações;
- processamento automatizado de informações da Web por meio de agentes inteligentes de software, migrando-se de um ambiente legível por humanos para uma rede mundial de processos compatíveis, com base no RDF;

O RDF foi desenvolvido para representar a informação com o mínimo de limitações, e pode ser utilizado em aplicações isoladas onde há formatos específicos e complexos de design. O RDF faz com que as informações tornem-se mais valiosas por meio do compartilhamento, ou seja, elas podem tornar-se mais disponíveis e conseqüentemente mais bem utilizadas por um número maior de aplicações e usuários da Internet. O RDF foi desenhado para atender aos seguintes requisitos de design:

- um modelo de dados que seja simples de estruturar e usar;
- uso de semântica formal (da lógica matemática) e inferência comprovável;
- vocabulário extenso baseado em URIs;
- sintaxe baseada em XML;
- suporte ao uso de *datatypes* de esquemas XML;
- opção para os usuários fazerem considerações simples sobre qualquer coisa;
- expressão universal de fatos simples;

- base para a união de afirmações e concordâncias sobre qualquer assunto.

A estrutura implícita de uma expressão em RDF pode ser vista como um gráfico rotulado, ou seja, com identificação de cada elemento, que consiste em nós e arcos diretos entre eles e rotulados, que fazem *links* entre os nós, conforme a Figura 4, a seguir.



**Figura 4: Modelo de expressão em RDF (W3C, 2004)**

Cada arco de propriedade representa uma declaração ou afirmação da relação entre os nós ligados, composta de três partes distintas:

1. a propriedade que descreve alguma relação (também chamada de predicado);
2. um valor que é o sujeito da declaração feita;
3. um valor que é o objeto da declaração.

A direção do arco é significativa por que sempre aponta para o objeto da declaração. O objetivo de um gráfico em RDF é a conjunção de todas as declarações que ele contém, por meio de quaisquer relações possíveis de serem feitas entre os entes, como por exemplo contém/ não contém, é/não é, é parte de, advém de, leva a, etc.

### **2.2.3 XML**

O padrão XML foi desenvolvido por um grupo de trabalho do W3C, originalmente conhecido como o *SGML Editorial Review Board* em 1996. O trabalho desenvolvido considerou que o XML será a linguagem amplamente utilizada na Internet e dará suporte a uma variedade de aplicações. Ele deve ser compatível com o SGML, a linguagem antecessora que deu origem ao XML, e é considerada uma linguagem de fácil escrita para programas que processam documentos em XML. (W3C, 2006)

Para que um documento seja bem escrito em XML, alguns requisitos devem ser seguidos, conforme orientado pela recomendação do W3C *Extensible Markup Language XML 1.0* (W3C 2004). Para tanto, diversos elementos devem constar da sintaxe do documento, os quais devem ser declarados por meio de comandos específicos que explicitam informações sobre o documento em produção.

O XML é uma linguagem de marcação concebida para armazenar e transmitir dados. Sua codificação permite que se identifiquem objetos, elementos entidades contidas em um texto como título, parágrafo, negrito, tabela, etc. (BAX, 2001). Assim como o HTML, o XML é uma linguagem que provê total liberdade para o autor marcar o seu texto. Abaixo, um exemplo de arquivo xml em seu código fonte:

```
<?xml version="1.0"?>
<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:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/unnamed.owl#"
  xml:base="http://www.owl-ontologies.com/unnamed.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="Linguística_Class">
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >gerada apenas por questões linguísticas e é detectada quando
  determinados enunciados em condições já previstas apresentam problemas de
  escolha linguística ao receptor, gerando uma flutuação entre duas ou mais
  condições aceitáveis.
  Não se contamina com aspectos não linguísticos</rdfs:comment>
  <rdfs:subClassof>
```

A especificação 1.0 do XML, recomendação do W3C (2004), define os parâmetros de validade e formação adequada de sintaxe do documento, e é uma orientação clara de como se ter êxito na composição do documento XML, assim como orientar o usuário a formar regras de restrição na estrutura do documento. A validade de um documento XML requer que este siga as regras expressas na definição do tipo de documento (DTD), a qual provê uma gramática livre de contexto para um tipo de documento. As DTDs são gramáticas de regras para verificação dos tipos de elementos que compõem a estrutura de um documento, como capítulos, títulos, cabeçalhos, parágrafos, seções etc. (SOUZA, 2004).

Em alguns casos, as aplicações precisam de definições ou construtos de marcação que sejam mais informativos, ou regras de restrição mais rígidas, o que levou ao desenvolvimento, por parte do grupo de trabalho em XML do W3C, dessa recomendação.

Conforme apresentado nesta tese, há diversas especificações que descrevem a forma como documentos em XML devem ser criados e processados na Web Semântica. Em ambientes de autoria, muitas aplicações dependem do uso de uma ou mais dessas especificações de requisitos e deve-se considerar como a implementação dessas especificações interagem, e questões importantes como interoperabilidade surgem. Os requisitos e casos de uso descritos nessa recomendação expressam como a interação entre os documentos XML deve ocorrer (*XML Processing Model Requirements and Use Cases* do W3C, 2006). A recomendação apresenta um modelo conceitual da interação entre processos em XML, a linguagem para a descrição dessas interações, assim como as entradas e saídas de todo o processo. A recomendação apresenta os requisitos de como se deve configurar os parâmetros e o fluxo necessário para a composição de documentos em xml.

### **2.3 Ontologias**

Originalmente oriundo da filosofia, o termo ontologia foi enunciado há milênios por Aristóteles para designar a ciência das coisas de acordo com a sua essência. Hoje o termo refere-se também no contexto da representação da informação e da Inteligência Artificial como uma “*especificação explícita de uma conceituação*” (GRUBER, 1996). Para GUARINO (1998), a ontologia refere-se a um “*artefato constituído por um vocabulário usado para descrever certa realidade, mais um conjunto de fatos explícitos e aceitos que dizem respeito ao sentido pretendido para as palavras do vocabulário*”.

Uma vez que os seres humanos raciocinam de forma semântica, ou seja, a partir dos conceitos e relações entre os objetos, o mesmo poderia acontecer com os computadores. A Inteligência Artificial estuda esse fenômeno e tem apresentado soluções que comprovam essa possibilidade. Além de inventar robôs, os cientistas têm atuado no campo da representação do conhecimento por meio da aplicação de características humanas no processamento das informações eletrônicas, via ontologias.

A tentativa de associar conceitos em um ambiente eletrônico da mesma forma com que é feito no cérebro humano é o grande desafio da representação do conhecimento. Para tanto, as ontologias revelam-se numa solução promissora para

fazer com que os computadores representem e encontrem informações de uma forma intuitiva, como fazem os humanos.

Ontologias descrevem o discurso de uma área com base na linguagem formal da lógica matemática. Atualmente a pesquisa na área dá-se principalmente dentro de laboratórios dos cursos de pós-graduação em Matemática, Inteligência Artificial e Ciência da Informação das principais universidades do mundo. Para a Internet, as ontologias oferecem um entendimento compartilhado de um domínio de conhecimento, de modo a superar as diferenças que existem nas terminologias utilizadas em diferentes sistemas de conhecimento a fim de permitir a interoperabilidade semântica. “*Uma ontologia é uma lista finita de termos e as relações entre eles*” (ANTONIOU e van HARMELEN, 2004) Os termos em uma ontologia denotam conceitos importantes em um domínio. Esses conceitos são chamados de classes de objetos e possuem assim uma hierarquia de classes que são arquitetadas por um trabalho ardoroso de engenharia ontológica.

Podemos visualizar as ontologias como se fossem árvores conceituais cujas folhas são os termos e seus respectivos sinônimos, antônimos ou qualquer outra relação semântica que se queira atribuir.

As ontologias são compostas pelos seguintes elementos:

- classes (organizadas em uma taxonomia);
- relações (representam os tipos de interação entre os conceitos de um domínio);
- axiomas (usados para modelar sentenças sempre verdadeiras); e
- instâncias (utilizadas para representar elementos específicos, ou seja, os termos e conceitos) (ALMEIDA e BAX, 2003, GRUBER, 1996; NOY e GUINNESS, 2001).

Outros tipos de informação também estão presentes nas ontologias, como por exemplo propriedades – ação, verbo que descreve o tipo de relacionamento entre os conceitos; restrições de valor – tratam as exceções nas relações; regras de separação (*disjointness statements*) – destacam os valores de algumas classes; e finalmente a especificação de relações lógicas entre os objetos – número de objetos atribuídos às classes (ANTONIOU e van HARMELEN, 2004).

Segundo o W3C (W3C, 2004) uma ontologia define os termos utilizados para descrever e representar uma área de conhecimento, por isso elas podem ser usadas por pessoas, bases de dados e aplicações que precisam compartilhar

informações de domínios. Um domínio pode ser um assunto ou área específica de conhecimento como medicina, metalurgia, botânica, gastronomia, ou qualquer outra área. As ontologias incluem definições básicas de conceitos, legíveis por computador, e as relações entre eles, ou seja, as ontologias codificam o conhecimento de um domínio, tornando esse conhecimento reutilizável por outros sistemas ou pessoas.

O uso de ontologias é crítico para aplicações que buscam pesquisar e integrar informações de diversas comunidades e assuntos. Embora os esquemas e os DTDs do XML são suficientes para a troca de informação entre sistemas, a carência de informação semântica impede que as máquinas usem novos vocabulários em XML, uma vez que um mesmo termo pode ser usado com noção diferente em outros contextos, e termos diferentes podem ser usados por conceitos que possuem o mesmo significado. O RDF e seu esquema podem resolver esses problemas, ao permitir que relações semânticas simples sejam associadas entre si por meio de identificadores. Com o esquema RDF, pode-se definir classes que possuem múltiplas subclasses ou superclasses, e pode definir propriedades que por conseguinte podem apresentar sub-propriedades, domínios e abordagens. Desta forma, o esquema RDF é uma linguagem ontológica simples, no entanto, para alcançar a interoperabilidade entre diversos esquemas, uma correlação semântica mais incrementada torna-se necessária. Daí surgiu a linguagem OWL, hoje tida como padrão recomendado pelo W3C (2004) que complementa as limitações do XML e do RDF.

Desde o surgimento da idéia da Web Semântica, há dez anos apenas, grupos de especialistas têm desenvolvido ontologias e várias tecnologias relacionadas, como ferramentas para construção de ontologias, edição, integração, geração automática de ontologias etc. Diversos são os trabalhos feitos na última década em tecnologias de ontologia e pesquisas têm sido realizadas, demonstrando o estado da arte sobre metodologias, aplicações e projetos em ontologias (DENY, 2004; ANTONIOU 2004; ALMEIDA e BAX, 2003; GÓMEZ-PÉREZ, 2002).

### **2.3.1 Aplicabilidades das ontologias**

Destacamos, segundo ALMEIDA e BAX (2003) as seguintes aplicações que fazem uso de ontologias:

- projetos relacionados à gestão do conhecimento;
- projetos relacionados a comércio eletrônico;
- projetos relacionados ao processamento de linguagens naturais;
- projetos relacionados à recuperação da informação na Web;
- projetos relacionados a educação.

Quanto à tipologia de ontologias, os autores categorizam:

- ontologias de alto nível;
- ontologias lingüísticas;
- ontologias para empresas;
- ontologias para domínios específicos;
- repositórios de ontologias.

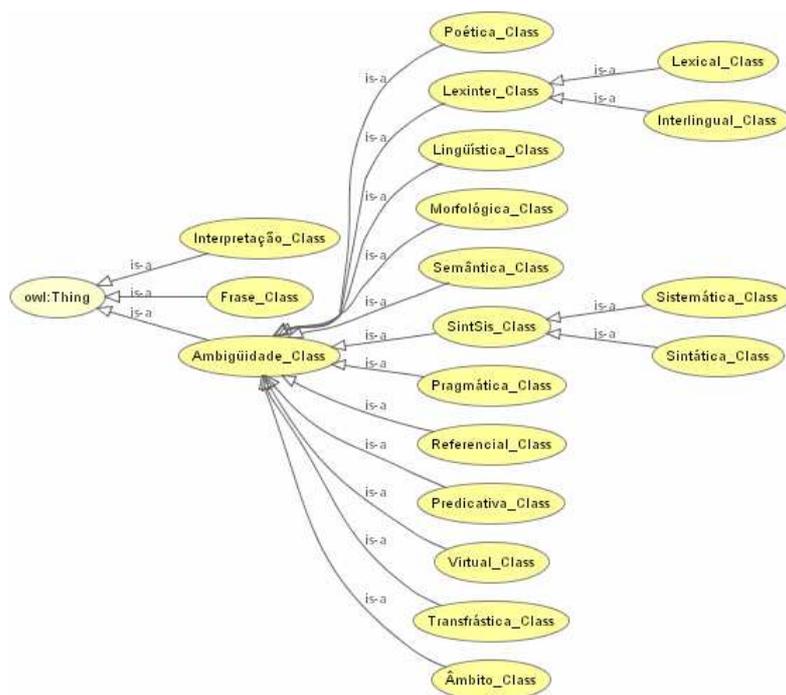
Com relação à engenharia ontológica, ressaltamos:

- metodologias para construção de ontologias;
- metodologias para construção de ontologias em grupo;
- metodologias para aprendizado sobre a estrutura de ontologias;
- metodologias para integração de ontologias;
- linguagens de construção de ontologias;
- métodos para avaliação de ontologias.

Algumas metodologias têm sido desenvolvidas com o propósito de gerenciamento de conhecimento com o apoio de ontologias como a CommonKADS (SCHREIBER et al, 1998), onde se tem as ontologias atuando como modelos de especialidades para o desenvolvimento de sistemas baseados em conhecimento; e SURE (2002) uma completa “*metodologia para o planejamento e desenvolvimento de ontologias baseadas nas aplicações da administração do conhecimento*”.

### **2.3.2 Visualização de ontologias**

As ontologias podem ser visualizadas de forma gráfica, conforme a Figura 5, a seguir.



**Figura 5: Exemplo de uma ontologia (SILVA, 2006)**

Pode também ser apresentada em forma de taxonomia (Figura 6).

Tabela 1: Taxonomia enumerativa das ambigüidades da Língua Portuguesa

	Classificação	Subclassificação
Ambigüidades	Âmbito	
	Interlingual	
	Lexinter	Lexical Interlingual
	Linguística	
	Morfológica	
	Poética	
	Pragmática	
	Predicativa	
	Referencial	
	Semântica	
	Sintsis	Sistemática Sintática
	Transfrástica	
	Virtual	

**Figura 6: Exemplo de Taxonomia (SILVA, 2006)**

Ou em forma de texto, o código fonte em XML (RDF e OWL), a seguir.

```

<?xml version="1.0"?>
<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:owl="http://www.w3.org/2002/07/owl#"
  xmlns="http://www.owl-ontologies.com/unnamed.owl#"
  xml:base="http://www.owl-ontologies.com/unnamed.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="Linguística_Class">
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >gerada apenas por questões linguísticas e é detectada quando
    determinados enunciados em condições já previstas apresentam problemas de
    escolha linguística ao receptor, gerando uma flutuação entre duas ou mais
    condições aceitáveis.
  Não se contamina com aspectos não linguísticos</rdfs:comment>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="Ambigüidade_Class"/>
    </rdfs:subClassOf>
    <owl:disjointWith>
      <owl:Class rdf:ID="Virtual_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Lexinter_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="SintSis_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Âmbito_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Transfrástica_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Poética_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Pragmática_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Semântica_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Referencial_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Predicativa_Class"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Morfológica_Class"/>
    </owl:disjointWith>
  </owl:Class>
  <owl:Class rdf:about="#Referencial_Class">
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >está relacionada a enunciados potencialmente ou efetivamente
    incompletos, possui a característica de ser uma ambigüidade elíptica. O
    efeito elíptico determina que um enunciado não pode conter todas as
    informações necessárias para o seu completo entendimento, logo é sempre
    possível acrescentar um elemento que possa, necessariamente contribuir
    para o seu entendimento causando um efeito de precisão,
    modificação ou também causando mais contradição ao enunciado em
    questão</rdfs:comment>
    <owl:disjointWith>

```

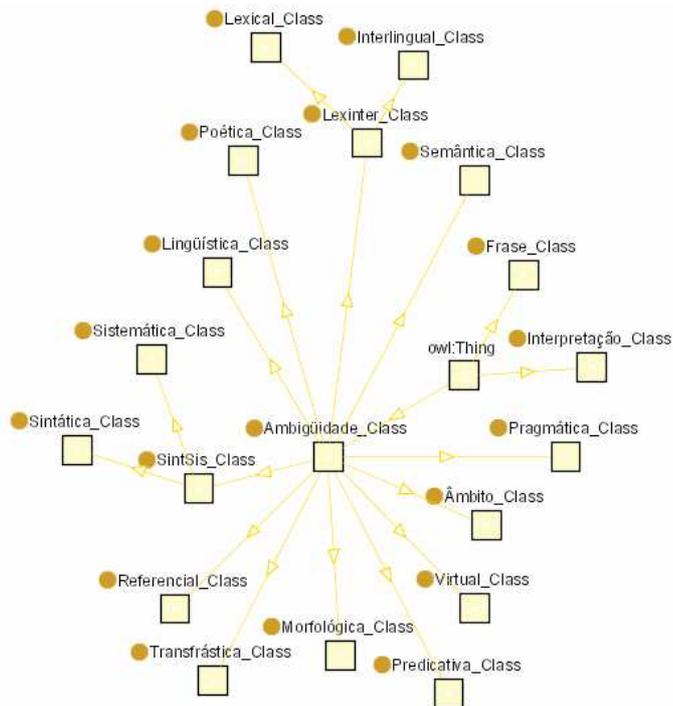


Figura 7: Visualização de ontologia (SILVA, 2006)

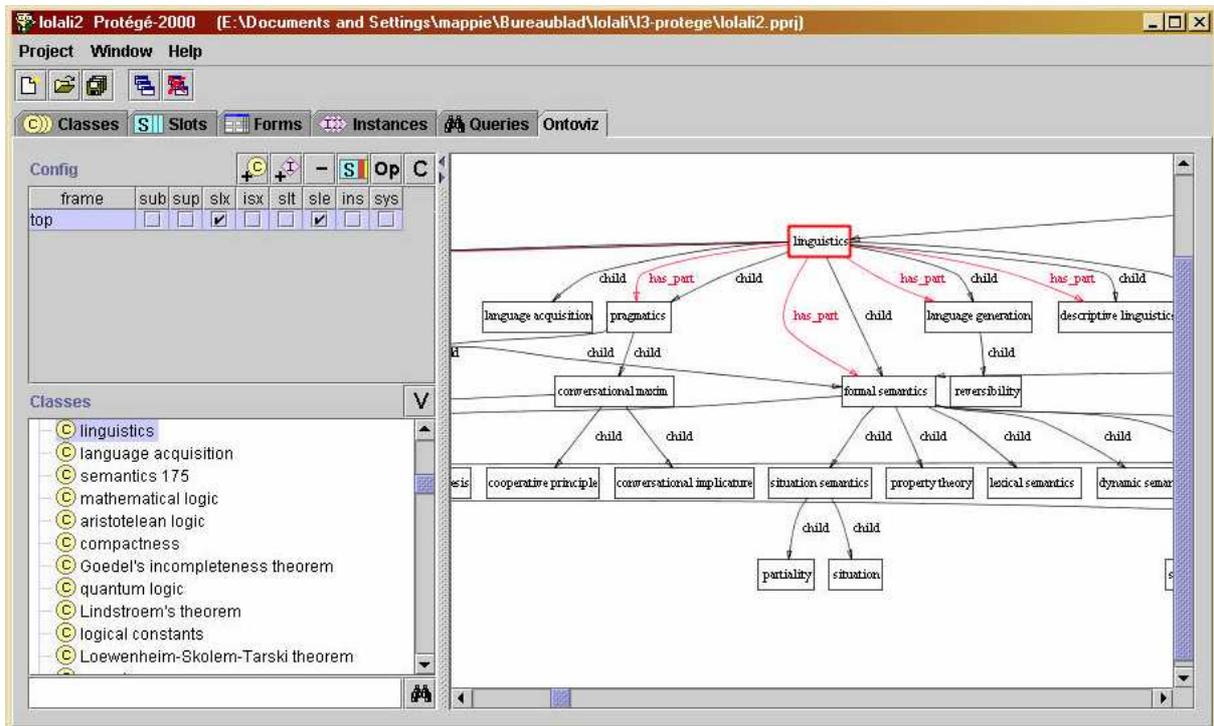


Figura 8 - Editor de Ontologias (Protégé) com visualização no OntoViz (van Schien, 2003)

Com ajuda das ontologias, a pesquisa e visualização da informação pode ser feita por meio de especializações ou generalizações de termos usados para identificar a informação, expandindo assim as possibilidades de formulação de *queries* de pesquisa em bancos de dados. Isso só é permitido por meio de relações lógicas numéricas, atribuindo assim uma linguagem formal que expressa o conhecimento e provê uma semântica formal bem compreendida com a ajuda de agentes inteligentes (robôs) que tornam o conhecimento tácito em explícito. Tal transformação acontece por meio de mecanismos de inferência que são deduzidos das formas de representação do conhecimento, é chamada de raciocínio ontológico de agentes inteligentes.

Concluimos assim que as ontologias são representações formais e especificações explícitas de uma conceituação (GRUBER, 1996). Com o advento da Web Semântica, diversos grupos de trabalho e pesquisa têm apresentado, junto ao W3C, um grande número de tecnologias baseadas em ontologias para uso e representação da informação já publicada. A questão que se faz quanto ao desenvolvimento de ferramentas e tecnologias para a Web Semântica é a de que elas devem compartilhar uma linguagem comum, e para isso o raciocínio ontológico deve ser aplicado nos processos humanos e eletrônicos. As linguagens da ontologia, como Shoe, DamL, têm sido adaptadas e transformadas no atual padrão do W3C, OWL (W3C, 2004), de modo a superar as limitações existentes no XML e RDF(S). O uso de ontologias para dar suporte à troca de informações é hoje o principal foco dos desenvolvedores de tecnologias para Web Semântica, que buscam ter seus serviços interoperáveis a essa estrutura conceitual projetada por Berners-Lee.

O sucesso do uso de ontologias está consagrado por meio da aplicabilidade de soluções de busca e localização de informação, integração, *querying*, apresentação, navegação e visualização da informação (FLUIT, 2003; STUCKENSCHMIDT, 2003). Podemos pensar em como esse aparato ontológico pode ser usado para a produção de novos conhecimentos no mundo da publicação eletrônica. Para HENDLER (2002), as tarefas de anotação, *link* entre páginas Web e ontologias, criação, avaliação e integração de ontologias deve ser o mais automático ou semi-automático possível, para não sobrecarregar o autor.

O estado da arte revela a tendência de se usarem ontologias como base para a representação do conhecimento, compartilhamento, gestão, modelagem, engenharia, educação, dentre outras. Se usarmos o mesmo *rationale* para a tarefa de autoria, concluímos que as ontologias são um recurso potencial para a produção de conhecimento para a Web Semântica.

Destarte, esta pesquisa busca respostas para a questão de se usarem ontologias para dar suporte à produção de documentos, permitindo que o autor esteja consciente das estruturas conceituais do assunto sobre o qual está escrevendo, oferecendo ao autor uma linguagem comum que permite localizar, acessar, recuperar, reutilizar e disseminar informações de uma forma multilinear, controlada e não ambígua.

#### **2.4 Tecnologias baseadas em ontologia para a Web Semântica**

Nesta seção, mostraremos como a Web Semântica de Tim Berners-Lee é um universo eletrônico promissor onde o conhecimento será criado e conectado de um modo semântico revolucionário, confirmando as visões de Vannevar Bush, Douglas Englebart e Ted Nelson. Isso mudará completamente a forma com que documentos estão sendo produzidos e disponibilizados na Web, com o crescimento de tecnologias de ontologias que permitem a anotação, representação, navegação e visualização semântica da informação, dentre outras funcionalidades.

A Web foi originalmente criada para se tornar um ambiente de informação no qual usuários não apenas navegassem pelos hipertextos, mas que eles também pudessem editar e produzir novos conteúdos e informações sem barreiras (BERNERS-LEE, 1999). Essa foi a idéia original do primeiro editor e navegador criado pelo Tim Berners-Lee (QUINT, 2005). No entanto, com o lançamento do software Mosaic, que pela primeira vez permitiu a visualização de informação orientada a caracter, esses dois mundos foram separados: o de escrita e o de pesquisa. Muito embora exista um grande número de métodos disponíveis e amplamente utilizados para geração de conteúdo da Web em geral, ainda há uma grande necessidade de desenvolvimento na área de autoria de documentos para a Web Semântica. Algumas ferramentas de anotação e autoria estão sendo desenvolvidas para a Web Semântica visando integrar o ambiente de navegação como o de edição, por meio da anotação semântica, aumentando assim a capacidade de representação do conhecimento.

Buscamos na literatura referências a esse tipo de abordagem: das 13 ferramentas listadas no site *Semantic Web Annotation and Authoring* (SEMANTIC, 2006), 10 ferramentas são utilizadas exclusivamente para anotação semântica, e três oferecem funcionalidades de autoria. São elas: *Semantic Word* (TALLIS, 2003), *Ont-o-Mat* (HANDSCHUH, 2002) e *Smore* (KALYANPUR, 2005). Ao estender a pesquisa a outras fontes de informação, identificamos três outras ferramentas que também lidam com autoria ontológica, são elas: *Cohse* (CARR, 2001), *WickOffice* (CARR, 2004) e *Trellis* (GIL, 2002). Essas seis ferramentas foram analisadas, no Capítulo 3, a partir de um conjunto de características sistematizadas.

Neste capítulo falaremos sobre algumas ferramentas que estão sendo desenvolvidas no âmbito da Web Semântica, divididas em cinco grandes grupos funcionais:

- anotação semântica de documentos;
- geração de metadados;
- *links* entre termos de documentos e ontologias externas;
- explicitação da estrutura retórica do documento para revisão de pares.

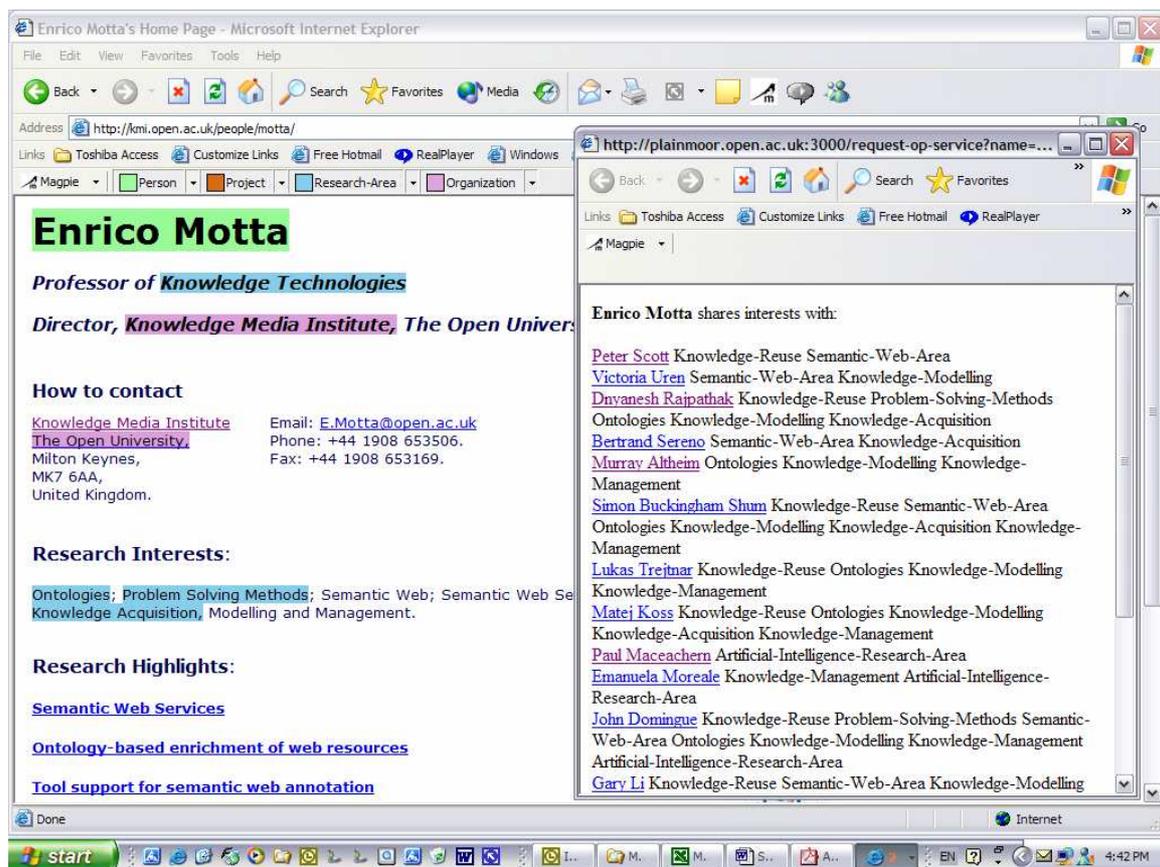
#### **2.4.1 Anotação Semântica de Documentos**

Anotação semântica é a tarefa de se criar um *link* entre um objeto no texto ao seu descritor semântico, por meio de um mecanismo de marcação visual. No contexto da Web Semântica, a anotação semântica pode ser feita com apoio das ontologias.

Diversas ferramentas foram desenvolvidas para fazer anotação semântica de documentos para propósitos específicos (SEMANTIC, 2005; FAABORG, 2003, ALMEIDA e BAX, 2003), como por exemplo, para interpretação de conteúdos de documentos. A anotação semântica ocorre da seguinte forma: uma ontologia é associada ao documento pelo usuário e os termos presentes no documento reconhecidos pela ontologia são coloridos e transformados em *links* para outros recursos de conhecimento semanticamente conectados. Assim, é criada uma camada semântica para navegação conceitual em forma de hipertexto.

Magpie (MOTTA, 2003) é um exemplo dessas aplicações, e foi desenhada para dar apoio à navegação semântica, ou seja, permitir a interpretação de documentos da Web, por meio de uma camada semântica ontológica. Por meio do

reconhecimento de descritores de informações presentes nas ontologias, o sistema torna as páginas visitadas em hipertextos interligados por “*hiperlinks ontológicos*”, montando assim uma base de dados inter-relacionada (Figura 9). Isto se dá por meio de NER (*named-entity recognition*), anotação e representação formal da ontologia, provendo assim possibilidades de navegação conceitual e aquisição do conhecimento.



**Figura 9 – Tela do Sistema Magpie**

O Magpie difere-se de outras ferramentas de anotação tradicionais pois separa a marcação das páginas, permitindo uma anotação mais dinâmica. Estudos indicam que as tecnologias de anotação semântica (ANGELE, 2001) tendem a fazer a anotação nas bases de dados e não nos documentos em si, visto que as páginas da Web são dinâmicas, mudam constantemente e com isso podem perder os marcadores (HANDSCHUH, 2002).

Outro exemplo de ferramenta com os mesmos propósitos é a KIM (POPOV, 2004), uma plataforma tecnológica de anotação semântica da informação, indexação e recuperação da informação por meio de ontologias e bases de

conhecimento. A extração também ocorre via NER a partir de uma ontologia específica que faz um *link* entre os termos no texto com seus tipos referentes, por meio das relações de atributos e propriedades dos conceitos.

Essas duas ferramentas se destacam no rol de ferramentas similares, pelo fato de oferecerem mais características e funcionalidades, agregando valores como a indexação e recuperação da informação como resultado da representação realizada automaticamente pela anotação semântica baseada em ontologias.

A extração e marcação da informação ajudam na representação automática de objetos, pelo reconhecimento de conceitos presentes em textos e automaticamente gerando formas de representá-los para análise e recuperação da informação. *Gate* (BONTCHEVA, 2004) também é um sistema de extração da informação para o gerenciamento de conteúdo e anotação semântica, que permite a visualização de ontologias relacionadas a fontes em um documento (Figura 10).

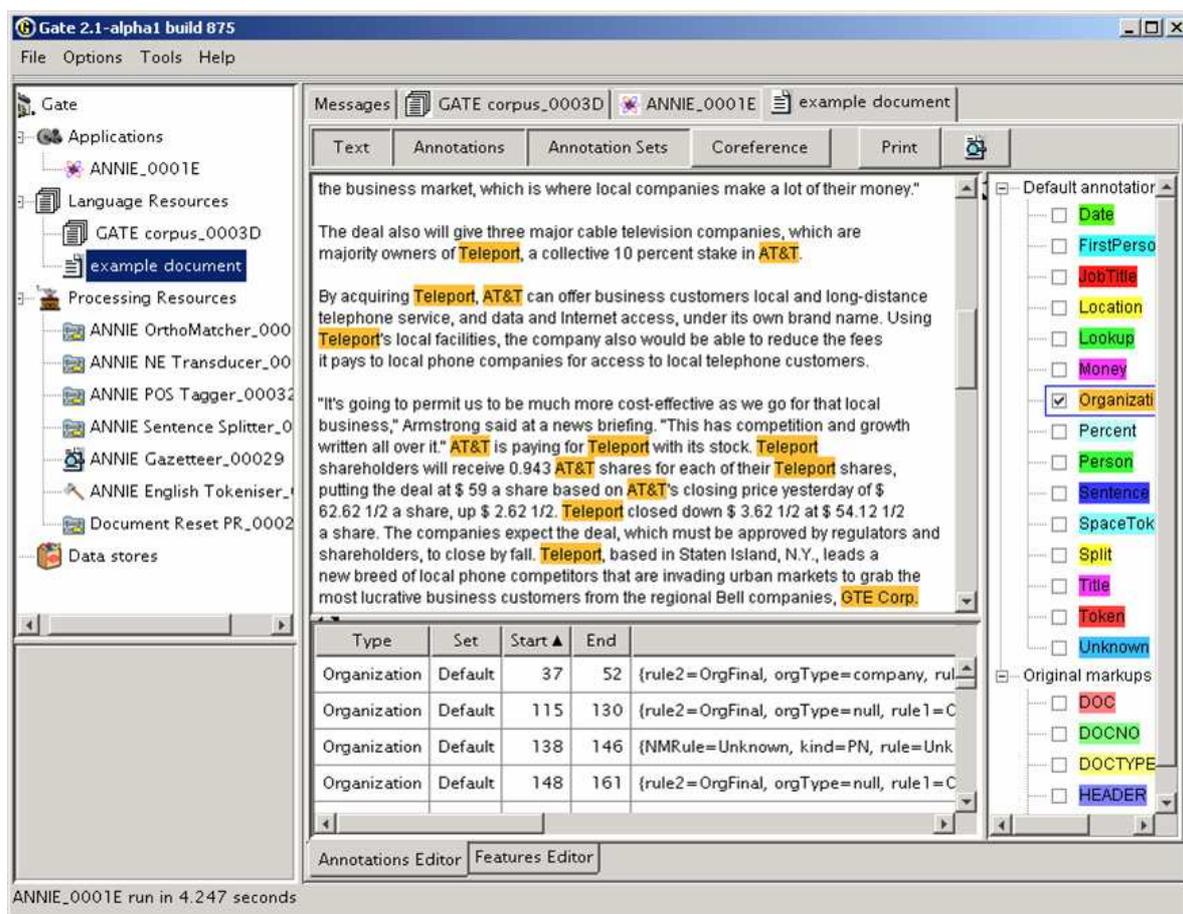


Figura 10: Tela do Sistema Gate

Outras ferramentas também são usadas para o processamento de linguagem natural, como o *Ellogon* (MARKANTONATON, 2002) uma plataforma multilíngüe de engenharia de texto baseada em processamento de linguagem natural (PNL). Ela é usada para pesquisa em sistemas de PLN e sistemas de engenharia lingüística.

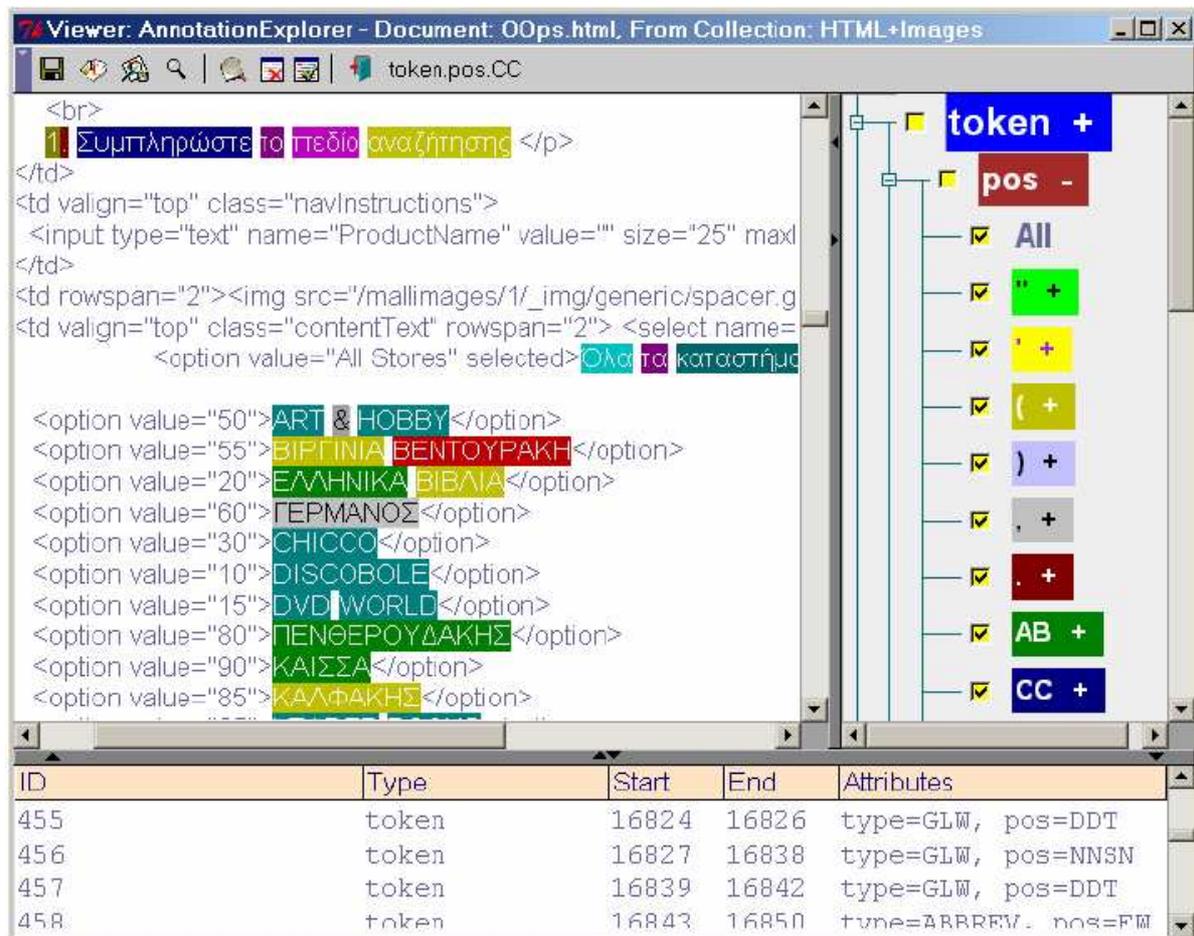


Figura 11- Tela do Sistema Ellogon

#### 2.4.2 Geração de Metadados

Os metadados, conhecidos como informações estruturadas sobre dados, ou dados sobre dados, são referências criadas para ajudar na identificação da informação eletrônica, por meio de rótulos categóricos de informação. Geralmente são gerados apenas depois que um documento está pronto. Veremos que na Web Semântica, o estado da arte revela a busca por mecanismos que auxiliem na construção de metadados durante a produção do documento.

Há modelos consagrados de metadados disponíveis para uso, e constituem-se em padrões internacionais de comunicação de informações bibliográficas legíveis

por computador. Na Web Semântica é recomendado o uso o Dublin Core (DUBLIN, 2004), que desde 1995 tem sido o modelo mais utilizado para representar documentos em ambientes eletrônicos ou não. Ele é um vocabulário controlado criado a partir da necessidade de se separar a busca da informação do meio em que está armazenada. O Dublin Core baseou-se no Modelo MARC (LOC, 2005) desenvolvido na Biblioteca do Congresso Americano em Washington, e também é bastante utilizado no mundo inteiro, especialmente por bibliotecas, digitais ou não.

Os elementos que compõem o padrão de metadados do Dublin Core são:

1 Título	
2 Criador	9 Formato
3 Assunto	10 Identificador
4 Descrição	11 Fonte
5 Publicador	12 Idioma
6 Colaborador	13 Relação
7 Data	14 Cobertura
8 Tipo	15 Direitos

Na Internet, os serviços de busca recuperam grandes quantidades de informações que não são pertinentes ou não satisfazem as necessidades do pesquisador. Isso porque os documentos disponíveis na rede carecem de dados suficientes para a sua descrição e representação. Os metadados cumprem a função de identificar um recurso em meios eletrônicos, como por exemplo (IANNELA, 1997) resumir o significado dos dados, permitir busca, determinar relevância do dado, recuperar e usar a cópia de um dado, instruções de interpretação do dado, informações sobre copyright, validade, propriedade e criação, relação com outros recursos, controle de gestão, etc. Após avaliar as aplicações dos metadados e as definições canônicas de metadados, SENSO (2003) chega a uma definição própria, abrangente e que inclui tendências do uso de metadados:

*“metadado é toda aquela informação descritiva sobre o contexto, qualidade, condição ou características de um recurso, a partir de um objeto que tem a finalidade de facilitar a sua recuperação, autenticação, avaliação, preservação ou interoperabilidade”.*

Um novo modelo de produção automática ou semi-automática de metadados pode ser encontrado em algumas tecnologias de que temos conhecimento. Waard (2003) questiona *“até que ponto um autor é capaz, pode ser capaz e está querendo estruturar a contribuição”* de um sistema de autoria, considerando que o metadado é

um elemento crucial para aumentar a recuperação depois que um documento é publicado.

O estado da arte revela que para a Web Semântica diversas soluções estão sendo apresentadas. STUCKENSCHIMIDT (2003) apresenta formas de geração e gerenciamento de metadado para sistemas de informação baseados na Web. O objetivo é permitir o menor nível de iteração entre usuário e sistema, permitindo uma geração automática dos metadados. Sistemas como o CREAM propõem soluções para a criação de metadado durante e depois da publicação (HANDSCHUH, 2002). SEELING (2003) propõe a união de tecnologias de *datamining* para aumentar a capacidade de recuperação de informação, explorar metadados para navegação baseada em ontologias para documentos não estruturados. OLESCHUK (2003) propõe uma metodologia para a comparação de similaridades semânticas dentro dos documentos.

Waard (2003) apresenta as categorias possíveis dos tipos de metadados para que se considere quem é responsável pela criação, validação e direitos autorais desses elementos. Identifica os atores responsáveis no processo de publicação eletrônica, associando-os a suas respectivas tarefas. A proposta apresentada é a de se tratar de uma forma mais automática possível a geração de metadados, sem sobrecarregar o autor. Essa abordagem apresenta as categorias possíveis dos tipos de metadados para que se considere quem é responsável pela criação, validação e direitos autorais, identificando os atores responsáveis pelo processo de publicação eletrônica e associando-os a suas respectivas tarefas. Waard (2003) divide tipos de metadados em três, com respectivas funções em ambiente eletrônico: conteúdo (para interpretar e validar); localização (para localizar e conectar ou interpretar e validar); formato (para manipular).

#### **2.4.3 Links entre termos de documentos e ontologias externas**

Na Web Semântica, a escrita de textos deixa de ser uma tarefa simples de registro de termos no papel, e torna-se principalmente uma atividade de edição semântica desses termos, de forma a relacioná-los com outros termos e documentos, por meio de códigos embutidos que são lidos por outros programas, convertidos em muitos formatos de representação. Assim, consideramos que a produção de textos e documentos para a Web Semântica é uma forma de autoria e não apenas de escrita.

Nesta seção veremos algumas pesquisas científicas de sistemas que visam associar os documentos a ontologias. Em princípio, consideramos as visões de KIM (2002) de que “*as ontologias serão amplamente utilizadas, se existirem ferramentas de desenvolvimento de ontologias as quais possam ser usadas por profissionais do conhecimento, não necessariamente por ontologistas*” (p.10). De fato, uma ferramenta de ontologia precisa ser avaliada de acordo com a “*facilidade de uso e capacidade de expressar ricos conceitos sem expertise na representação de conhecimentos complexos*”.

A segunda previsão revela que “*ontologias são previstas para serem usadas amplamente, se uma ontologia desenvolvida por um profissional do conhecimento é útil ao profissional, independentemente de ser ou não usada para compartilhamento de dados. Assim sendo, as ontologias poderão ser largamente adotadas primeiramente para especificação de softwares*” (KIM, 2002, p.10). Ele afirma ainda que os usuários não contribuirão para popular as bases de dados ontológicas caso seja uma tarefa muito difícil e demandar muito tempo, sobrecarregando suas tarefas. Isso se refere à representação da informação que precisa estar associada a tarefas de conhecimento comum que os usuários usam, como, por exemplo, autoria de textos e pesquisa.

Uma proposta tecnológica avançada, que se destaca na tarefa de anotação semântica, é o Projeto Cohse (CARR, 2001), desenvolvido pela Universidade de Southampton. O projeto teve como objetivo implementar uma hipermídia baseada em ontologia que representasse modelos conceituais de termos de documentos e suas relações por meio de um serviço de raciocínio ontológico. Essa estrutura provê *links* entre os documentos via metadados que incrementam a consistência e alcance dos *links* dos documentos da Web, tanto durante a recuperação da informação quanto durante a autoria. Segundo CARR (2001), os conceitos presentes nos documentos podem ser ligados durante o momento da autoria, a fim de que outros documentos que tenham os mesmos conceitos possam ser interligados entre si para permitir uma navegação semântica. Este projeto inspirou o grupo que desenvolveu o Magpie (MOTTA, 2003).

O Cohse foi desenhado para permitir navegação conceitual de documentos em um ambiente de descoberta que está baseado em ontologias, provendo i) um serviço de raciocínio ontológico que representa modelos conceituais de termos do documento e suas relações; ii) um serviço de *link* baseado em Web para hipermídia,

que faz o *link* entre os termos. Essas duas características estão integradas formando um sistema de hipermídia que faz com que documentos sejam ligados entre si por meio de metadados que descrevem os seus conteúdos, e assim melhorando a consistência e abrangência dos *links* de um documento na Web tanto durante o momento da recuperação quanto no momento da autoria (CARR, 2001).

Uma vez que conceitos de documentos são ligados entre si durante a autoria, outros documentos associados com os mesmos conceitos tornam-se conectados um com os outros. Nesse sistema, os autores compartilharam da necessidade dos metadados serem mais descritivos para assim melhorar o processo de autoria, assim como fonte de pesquisa para leitores.

O sistema Cohse integra uma terminologia de domínio específica, construída por relações lógicas, dando suporte à classificação dinâmica de termos, gerenciamento automático de terminologias e catalogação flexível de documento durante autoria e navegação de documentos da Web.

Conforme pode ser visto na tela a seguir (Figura 12), os documentos, quando são interligados no momento da autoria, tornam-se associados a outros documentos que apresentam conceitos correlatos.

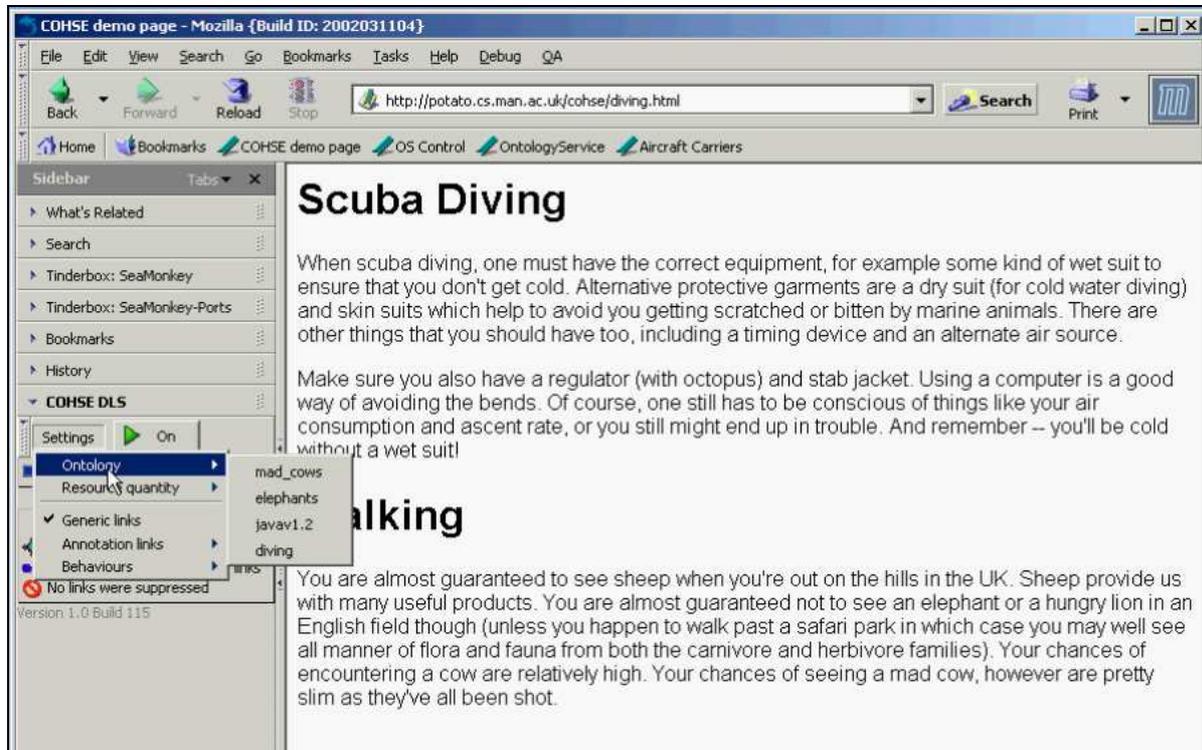


Figura 12 – Tela do Sistema Cohse

O exemplo de outra ferramenta com essas funcionalidades, apresentado por HAYASHI et al (1998), introduz uma nova abordagem do uso de ontologia como fonte de conhecimento para ajudar autores a construir modelos para sistemas educacionais. O autores apresentam a ferramenta *SmartTrainer/AT*, a qual é baseada em ontologia para sistemas de treinamento, construída sobre os seguintes requisitos: i) prover itens amigáveis para que os usuários possam facilmente usá-los para descrever suas próprias idéias sobre um IES (sistema educacional inteligente); ii) dar orientação apropriada para autores, baseada em princípios estabelecidos na tarefa de educar; iii) mostrar o comportamento dinâmico do IES em um nível conceitual para a averiguação de sua validade.

ALBERINK (2003) descreve o projeto do sistema *Topia*, em que usam tecnologias de cluster para gerar discurso a partir de relações semânticas e permitir a geração hierárquica e seqüencial de um discurso. Outro exemplo de ambiente com estas características de interligar autoria e ontologia, é o proposto por QUAN (2003), uma plataforma para orientar desenvolvedores de aplicações para a Web Semântica, chamada *Haystack* (QUAN, 2003), que não é uma ferramenta de autoria em si, mas provê um ambiente de desenvolvimento baseado em RDF, auxiliando os desenvolvedores a construir relações semânticas entre os conceitos utilizados nos sistemas.

Outras ferramentas foram identificadas nesses estudos como práticas potenciais na associação entre documentos e ontologias (NACK, 2002; ANDREASEN, 2004; VELARDI, 2001; AROYO 2002 e CRISTEA, 2004). Podemos concluir que o estado da arte revela diversas possibilidades e mostram que as ontologias, no âmbito documentário, podem ter diversas funcionalidades na representação do conhecimento.

#### **2.4.4 Explicitando a estrutura retórica do documento para revisão de pares**

Nos ambientes de autoria, a revisão e discussão da literatura é uma tarefa fundamental que permite a crítica de pares de uma mesma área, sobre um determinado artigo ou obra, para que se julgue a qualidade do trabalho, para fins de publicação. O estado da arte revela que essa tarefa pode ser feita com base em ontologias. *Claimaker* (UREN, 2003) é uma ferramenta de colaboração para modelagem de documentos em um ambiente de revisão de pares para argumentação, unindo visualização da informação com engenharia do

conhecimento. Nessa ferramenta, um ou mais revisores ou pares podem fazer seus comentários sobre documentos, conjuntos de dados e ferramentas, permitindo que os pesquisadores construam estruturas argumentativas (*claims*) para descrever e debater a contribuição da obra e a relação desta com a literatura correlata. As argumentações dos revisores são conectadas a versões em linha do documento, de modo que as argumentações passam a cumprir uma função de representação de informações (Figura 13). Por meio dessa malha, são construídas espécies de “ontologias de discurso” que explicitam uma linguagem comum para a construção dos modelos de argumentação.

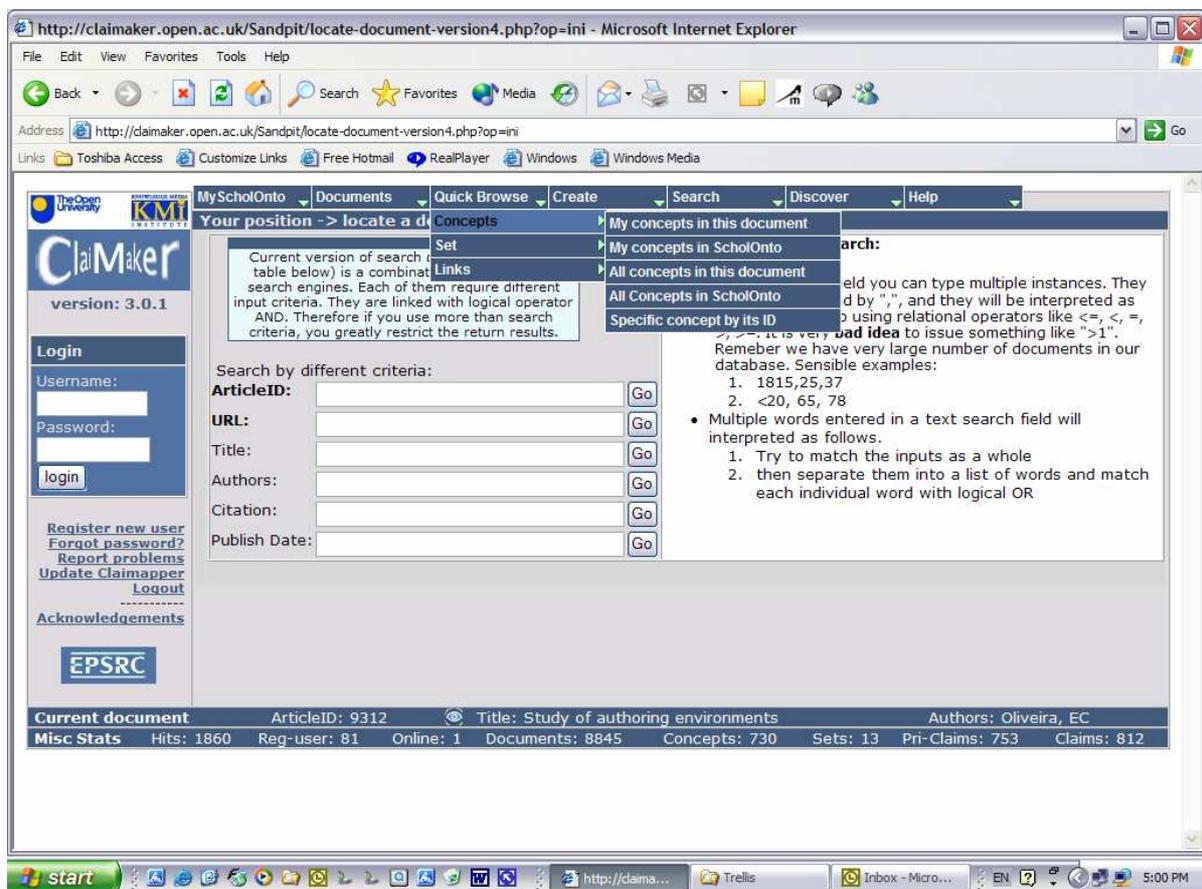


Figura 13 – Tela do Sistema Claimaker

Um outro sistema, que também visa interligar estruturas argumentativas é o *Trellis* (GIL, 2003). Ele oferece uma infra-estrutura para usuários escreverem seus próprios comentários sobre recursos de interesse, fazendo anotações de suas visões e opiniões, assim como qualificando e justificando essas considerações para a tomada de decisões. Isto se dá por meio de um editor (Figura 14) que permite que as argumentações e comentários inseridos pelos revisores possam corroborar ou refutar as idéias contidas no documento.

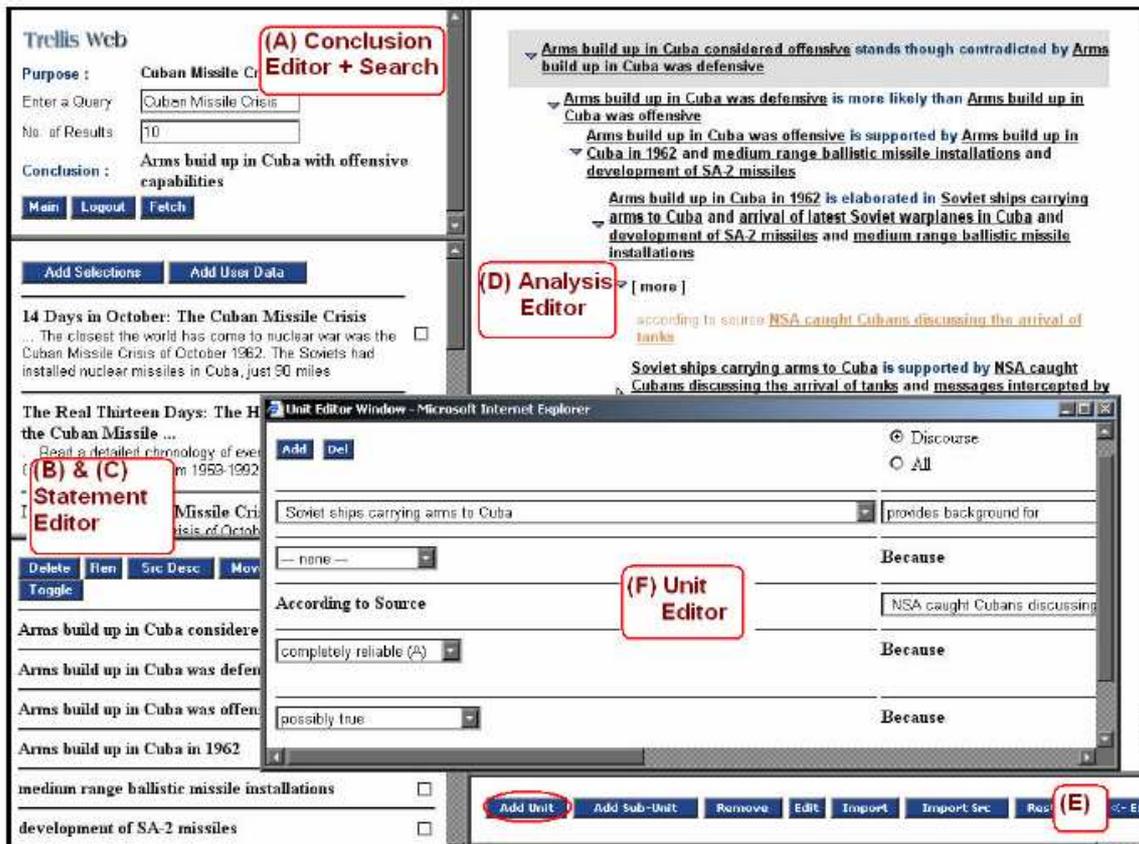


Figura 14 – Tela do Sistema Trellis

Um outro exemplo é o *D3E – Digital Document Discourse Environment* (SUMNER, 1998) – um sistema experimental que oferece suporte à publicação de documentos na Web, por meio de facilidades de discurso integradas e componentes interativos com o objetivo de prover ambientes de produção de documentos a partir de uma abordagem centrada no discurso dos participantes. CARACIOLLO (2002) também apresentou um sistema de acesso à informação baseado em tópicos, por meio de tecnologias de recuperação da informação que une o texto com uma hierarquia de conceitos com base em lógica descritiva. O ambiente oferece uma forma de explicitação da retórica de um documento, mas neste caso para facilitar a navegação conceitual. Consta de um arquitetura para disseminação eletrônica de manuais científicos, que padroniza a forma com que uma publicação eletrônica deve ser apresentada, a partir de um menu conceitual, e facilidades de integração com outros recursos como visualizadores e simuladores, filmes e outras ferramentas.

Os sistemas mencionados acima representam um passo significativo no âmbito da autoria, uma vez que eles oferecem uma infra-estrutura para pesquisadores navegarem visualmente pelas redes de conceitos que são “ligadas” aos documentos-fonte. É como se fosse uma camada de conhecimento colocada

sobre a tarefa de revisão, tornando as argumentações em estruturas de navegação com base nos comentários feitos pelos revisores do documento. Essas ferramentas também podem ser usadas para modelar informações sobre literaturas de um assunto para análise da informação, pesquisa e ensino. No entanto, podemos pensar em como essas estruturas podem ser direcionadas para o processo e autoria, seja ela individual ou colaborativa.

#### **2.4.5 Conclusões**

A Web Semântica oferece um novo paradigma para a produção, representação, disponibilização, recuperação e uso da informação. Esse paradigma é apoiado por linguagens específicas de comunicação e de formatação de arquivos, e uma representação semântica da informação por meio de ontologias. As novas tecnologias de ontologia sugerem um conjunto de possibilidades inéditas de representação da informação para a Web Semântica.

Com essa revisão de literatura, notamos que os sistemas citados provêm diferentes características e componentes para um ambiente de autoria baseado em ontologia. Nossa proposta é unir essas características e apresentar alguns requisitos para um ambiente de autoria que leve em conta o uso de ontologias e tecnologias relacionadas para melhorar as tarefas de redação e representação da informação. O objetivo deste trabalho é desenhar uma plataforma que possa integrar as características identificadas e compor um novo ambiente de autoria. Essa proposta sugere um ambiente de produção de conhecimento baseado em ontologia, integrado pelas tecnologias disponíveis em uma plataforma que permita aos usuários (autores, revisores, editores, etc.) executarem atividades necessárias durante a produção do conhecimento.

A seguir apresentamos de forma sucinta as funcionalidades e características de ambientes baseados em ontologia para Web semântica. Esse trabalho foi apresentado na conferência EIPub 2004, e publicado em OLIVEIRA (2004).

- **Autoria** – integração das tecnologias de edição e processamento de texto aos recursos de informação relacionados, às bases de conhecimento e às ontologias, de modo a oferecer aos autores um ambiente conectado por *hiperlinks* onde eles poderão facilmente navegar pelas fontes da informação e executar tarefas de autoria, automaticamente atribuindo ao seu documento a informação referenciada.

- **Anotação e marcação semântica** – anotação e marcação de termos relevantes que podem ser conectados por *hiperlinks* a outros documentos ou ontologias, de maneira automática ou semi-automática, provendo suporte para navegação conceitual, descoberta de conhecimento (*discovery*) e geração de metadado. A anotação pode dar apoio a autores, i) para reconhecimento de termos no documento que está *conectado* às ontologias, ii) para atribuição de elementos de metadados a conceitos no texto, iii) para prover uma navegação conceitual via *hiperlinks*. A anotação precisa ter o suporte de ontologias globais e de domínio, categorias gerais, servidores de índice, assim como por terminologias como tesouros e taxonomias, que podem sugerir ou restringir o uso de termos em um ambiente específico, seja pela padronização ou para evitar ambigüidades em uma comunicação especializada.
- **Navegação semântica** – visualização de conexões semânticas entre os conceitos e informações correlatas, por meio de uma navegação estruturada, e criação automática ou semi-automática de *hiperlinks* para as fontes de informação.
- **Geração de metadados** – atribuição de descritores de metadados para termos e frases dos documentos, automaticamente ou semi-automaticamente, para ser executado por cada usuário específico nessa tarefa (autor, editor, revisor). Os metadados podem ser criados a partir do termo ou do conceito que está registrado no documento, mas que também pode estar relacionado aos termos genéricos (TG) ou termos específicos (TE) de tesouros, taxonomias e ontologias.
- **Suporte multi-ontológico** – o uso de uma ou mais ontologias simultaneamente, no caso de assuntos que vão além de ontologias de domínio, ou são mais específicos do que aqueles presentes em ontologias globais. Considerando que o conhecimento pode ser gerado em múltiplos domínios, é fundamental que a ferramenta de autoria permita fazer *links* para mais de uma ontologia em um mesmo ambiente, e questões cruciais como integração e população de ontologias devem ser levadas em consideração.
- **Visualização da ontologia** – navegação semântica pelas ontologias de modo a permitir a aquisição de conhecimento (por máquinas ou humanos) e

a atribuição e localização de um conceito dentro de uma estrutura hierárquica aos seus conceitos relacionados para pesquisa e indexação. O estado da arte em visualização de ontologias revela ambientes de navegação conceitual onde usuários podem localizar o termo na ontologia e checar a disponibilidade de um conceito particular dentro do domínio.

- **Uso de termos e conceitos baseado em estrutura de *tesauros*** – suporte para o reconhecimento de sinônimos e antônimos dentro de uma estrutura conceitual, para sugerir o uso de termos preferidos, a fim de identificar e evitar variações terminológicas e ambigüidades, e assim melhorar a representação do conhecimento, atribuindo apenas o conceito específico do domínio para os termos relacionados no domínio.
- **Extração de informação** – reconhecimento automático de informação correlata presente em muitas fontes de conhecimento com a finalidade de parseamento e indexação da informação.
- **Engenharia de texto** – identificação de elementos semânticos em um texto para atividades de processamento de linguagem natural, tais como troca de informação, tradução, operabilidade com outros sistemas, assim como a visualização de árvores gramaticais (*parse trees*) que indicam as estruturas lingüísticas das frases.
- **Ambiente de descoberta de conhecimento** – pesquisa e navegação baseadas em ontologia pelas fontes em linha, para referência, citação, etc., fazendo marcação automática de fontes relacionadas e criação de *hiperlinks* do documento às paginas referenciadas.
- **Argumentação e revisão de pares** – uma estrutura para a revisão de pares com a proposta de criticar o artigo científico e julgá-lo, aceitando-o ou rejeitando-o, por uma comunidade científica de pares, baseados em ontologias compartilhadas para serem usadas pelos mesmos.
- **Publicação** – submissão de um documento às editoras, conferências, seminários etc., para a divulgação formal do conteúdo, baseado em padrões definidos por cada um deles. Modelos, diretrizes e recomendações orientam os autores a formatar a estrutura e os metadados dos documentos, seguindo as exigências e recomendações por eles providas.

## **2.5 Modelagem de dados em UML**

Nesta seção apresentaremos uma abordagem teórica e referencial da modelagem de dados adotada nesta tese para a especificação de alto nível do ambiente de autoria para a Web Semântica. A especificação de alto nível parte de dois pontos: a modelagem dos diagramas do sistema e ii) a especificação de requisitos de acordo com normas internacionais de engenharia de software.

Modelar é uma parte essencial de grande parte dos projetos de software, e é útil para projetos tanto médios como para os menores. Usando um modelo, aqueles responsáveis pelo sucesso de um projeto de desenvolvimento de software podem assegurar-se que a funcionalidade do negócio está completa e correta, as necessidades do usuário final são encontradas, e o projeto do programa suporta requisitos para escalabilidade, robustez, segurança, expansão, e outras características, antes que a implementação em código venha em retribuir mudanças difíceis e onerosas de se realizar (OBJECT, 2006).

A modelagem do ambiente proposto nesta tese teve como base os processos da UML – *Unified Modelling Language* – que é uma linguagem de diagramação ou notação para especificar, visualizar e documentar modelos de sistemas de software orientados a objeto. A UML não é um método de desenvolvimento, o que significa que ela não diz o que fazer primeiro e em seguida, ou como se desenhar o sistema, mas auxilia a visualizar o design e a comunicação entre objetos. A UML é controlada pela OMG (*Object Management Group*) e é um padrão da indústria para descrever softwares de maneira gráfica.

Para modelar um sistema, utilizamos os seguintes diagramas da UML.

### **2.5.1 Diagrama de caso de uso**

Na engenharia de software, um caso de uso (ou *use case*) é um tipo de classificador que representa uma unidade funcional coerente provida pelo sistema, subsistema, ou classe manifestada por seqüências de mensagens intercambiáveis entre os sistemas e por um ou mais atores. Pode ser representado por uma elipse contendo, internamente, o nome do caso de uso (WIKIPÉDIA, 2005).

Os casos de uso foram propostos inicialmente por Ivar Jacobson em sua metodologia de desenvolvimento de sistemas orientados a objetos OOSE. Posteriormente foi incorporado à UML tornando seu uso uma prática freqüente na

identificação de requisitos de um sistema. Nesse contexto um caso de uso descreve um comportamento que o software a ser desenvolvido apresentará quando estiver pronto.

Um software freqüentemente é um produto complexo, e sua descrição envolve a identificação e documentação de vários casos de uso, cada um deles descrevendo uma "fatia" que o software ou uma de suas partes deverá oferecer.

É importante notar que os casos de uso não descrevem como o software deverá ser construído, e sim, como ele deverá se comportar. Os casos de uso normalmente evitam o uso de termos técnicos, preferindo a linguagem de programação do usuário final. Normalmente os casos de uso são feitos tanto por quem desenvolve o software quanto pelos usuários que irão manipular esse software.

Os diagramas de caso de uso descrevem relacionamentos e dependências entre um grupo de caso de uso e os atores participantes no processo. É importante observar que diagramas de caso de uso não são adequados para representar o desenho, e não podem descrever os mecanismos internos de um sistema. Os diagramas de caso de uso são feitos para facilitar a comunicação com os futuros usuários do sistema, e com o cliente, e são especialmente úteis para determinar os recursos necessários que o sistema deve ter (KDE, 2005).

No diagrama de caso de uso, utilizamos a seguinte simbologia gráfica:

- a) desenho estilizado de um boneco para representar os atores do sistema;
- b) círculos ovais para representar as ações que os respectivos atores desempenham no sistema;
- c) linha contínua para associar ação ao ator que a realiza no sistema;
- d) linha tracejada para indicar ações adicionais realizadas pelos atores, que estão fora do sistema, mas que são incluídas como função pertinente.

### **2.5.2 Diagrama de Seqüência**

Em um sistema computacional orientado a objeto, os serviços (casos de uso) são fornecidos por meio da colaboração de grupos. Os objetos interagem por meio de comunicações de forma que juntos, cada um com suas responsabilidades, realizem os casos de uso (MACORATTI, 2005).

O diagrama de seqüência é uma ferramenta importante no projeto de sistemas orientados a objetos. Embora a elaboração dos diagramas possa consumir um tempo considerável para sistemas maiores ou mais complexos, eles oferecem as bases para a definição de uma boa parte do projeto como: os relacionamentos necessários entre as classes, métodos e atributos das classes e comportamento dinâmico dos objetos.

Um diagrama de seqüência é um diagrama de objetos, ou seja, ele contém como primitiva principal um conjunto de objetos de diferentes classes. O objetivo dos diagramas de seqüência é descrever as comunicações necessárias entre objetos para a realização dos processos em um sistema computacional. Os diagramas de seqüência têm esse nome porque descrevem ao longo de uma linha de tempo a seqüência de comunicações entre objetos. Como podem existir muitos processos em um sistema computacional, sugere-se proceder à construção dos diagramas de seqüência por caso de uso. Assim, toma-se separadamente cada caso de uso para a construção de seus diagramas de seqüência. De uma forma geral, para cada caso de uso constrói-se um diagrama de seqüência principal descrevendo as seqüências normais de comunicação entre objetos e diagramas complementares, descrevendo as seqüências alternativas e o tratamento de situações de erro (MACORATTI, 2005).

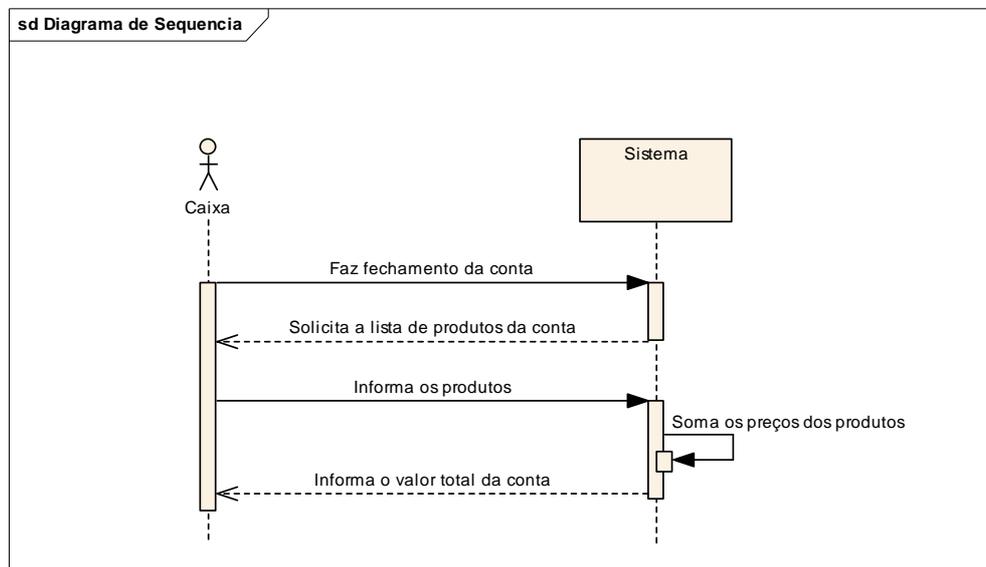
Utiliza-se também o termo cenário associado com os diagramas de seqüência. Um cenário é uma forma de ocorrência de um caso de uso. Como o processo de um caso de uso pode ser realizado de diferentes formas, para descrever a realização de um caso de uso pode ser necessário estudar vários cenários. Cada cenário pode ser descrito por um diagrama de seqüência. No exemplo do caso de uso "Cadastrar Aluno" de um sistema de controle acadêmico, podem-se considerar os cenários de inclusão, alteração e exclusão de aluno.

Um diagrama de seqüência do sistema ilustra eventos partindo de atores e estimulando o sistema. Dessa forma, os atores geram eventos para o sistema, pedindo que alguma operação seja feita.

Num diagrama de seqüência, cada objeto é apresentado com uma linha vertical que representa a "vida" do objeto. Acima dessa linha há uma caixa representando o objeto. Enquanto o objeto tem o controle (ou está aguardando para uma operação retornar o controle a ele), a linha de vida é uma caixa vertical. Nesse diagrama, o tempo vai de cima para baixo.

O ponto inicial de um diagrama de seqüência é, partindo-se de um caso de uso, identificar quais operações são efetuadas a partir de cada classe e em que ordem. A dimensão temporal é representada verticalmente.

Um exemplo de diagrama de seqüência pode ser desenhado assim:



Dessa forma, o diagrama acima será lido da seguinte forma.

1. O funcionário do caixa acessa o sistema e faz o fechamento da conta.
2. O sistema solicita a lista de produtos da conta para o funcionário do caixa.
3. O funcionário do caixa informa os produtos da conta para o sistema.
4. O sistema soma os preços dos produtos.
5. O sistema informa para o funcionário do caixa o valor total da conta.

### 2.5.3 Diagrama de Atividades

O diagrama de atividades descreve a seqüência das atividades de um sistema e mostra como uma atividade depende uma da outra. Um diagrama de atividade pode conter regiões denominadas *swimlanes*. Essas regiões são associadas a um objeto do modelo e, dessa forma, dentro de cada região, encontram-se as atividades relativas ao objeto da região. As atividades são conectadas por meio de arcos (transições), que mostram as dependências entre elas (SAUVÉ, 2006).

No diagrama de atividades, utilizamos a seguinte representação simbólica:

- a) círculos pequenos indicam o início e o fim do processo;
- b) caixas ovais representam as ações realizadas no sistema;

- c) caixas quadradas indicam os atores ou entidades externas que compõem o sistema;
- d) losangos deitados indicam uma decisão no sistema, que pode levar a uma ou mais ações alternativas conseqüentes;
- e) setas contínuas indicam relação de seqüência entre as atividades;
- f) setas tracejadas indicam relação de dependência a entidades ou ações externas;
- g) a barra indica o término de duas ou mais ações que juntas levam a uma ação conseqüente.

#### **2.5.4 Especificação de Requisitos do Sistema**

A especificação de requisitos de um sistema é uma atividade criteriosa que permite a explicitação de como um sistema é caracterizado para então poder ser desenvolvido. É fundamental que essa atividade seja feita de forma padronizada e com base em recomendações consagradas da engenharia de software. Para a especificação do sistema que iremos propor mais adiante nesta tese, utilizamos como base a norma *IEEE Std 830-1998 – IEEE Recommended Practice for Software Requirements Specification* (práticas recomendadas para a especificação de requisitos de software), desenvolvida pelo *Institute of Electrical and Electronics Engineers, Inc*, entidade baseada nos Estados Unidos e que se dedica à padronização de ambientes elétricos e eletrônicos, reconhecida mundialmente. Essa norma é utilizada amplamente para a elaboração de documentação que define a estrutura de requisitos para projetos de desenvolvimento de software no mundo inteiro (SAWYER, 2001).

A norma IEEE Std 830-1998 tem como objetivo apresentar um modelo de organização de requisitos de sistema, que possa ajudar no processo de criação, para que os clientes do sistema possam descrever precisamente o que deve ser desenvolvido, e para que os fornecedores de soluções possam compreender o que os clientes desejam.

Dentre os benefícios da especificação de requisitos por meio de uma estrutura organizada são: o estabelecimento de bases de acordo entre os clientes e fornecedores sobre o que o software deve conter; reduzir os esforços de desenvolvimento, visando à otimização do processo de produção; prover bases para

a estimativa de custos e de tempo de desenvolvimento da solução; prover uma infra-estrutura para validação e verificação da solução proposta, a partir de modelos de conformidade existentes ou sugeridos; facilitar a transferência do software de um ambiente para outro, seja tecnológico ou de usuário, permitindo uma maior flexibilidade à solução para um uso mais escalável; servir de base para a melhoria de todo o processo, uma vez que todo o trabalho estará formalizado e explicitado por meio de uma especificação que seja clara e objetiva.

A norma *IEEE Std 830-1998* está estruturada em cinco grandes itens (cláusulas). O primeiro item trata do escopo da norma, o segundo apresenta referências de outras normas utilizadas como base para a criação desta, o terceiro apresenta a definição de termos a serem utilizados, o quarto traz orientações básicas de como se escrever uma especificação de requisitos, e finalmente o quinto traz um detalhamento de cada uma das partes da especificação de requisitos.

As principais características de uma boa especificação de requisitos, conforme recomendado pela norma, são:

- i) correção – cada requisito deve estar representado por uma funcionalidade específica do software;
- ii) não-ambíguo – cada requisito apresentado deve ter apenas uma interpretação, e os termos utilizados devem ser listados em um glossário técnico para facilitar a compreensão;
- iii) completo – deve incluir todos os requisitos significativos, considerando também requisitos externos que possam influenciar na especificação do software. Deve também explicitar todas as respostas que o sistema apresentará, assim como referências a todas as figuras e citações presentes na especificação;
- iv) consistente – refere-se à consistência interna do documento de especificação para que cada requisito descrito não seja conflitante com outro existente;
- v) priorizado – refere-se à organização dos requisitos por grau de importância, necessidade, essencialidade ou criticidade dos mesmos;
- vi) verificável – os enunciados dos requisitos devem conter termos concretos e quantidades mensuráveis que possam permitir a interpretação da aplicabilidade dos requisitos;

- vii) modificável – que a estrutura e estilo utilizados para apresentar os requisitos possam ser alterados de modo fácil, completo e consistente, não deve ser redundante e deve apresentar os requisitos separadamente;
- viii) rastreável – a organização do documento deve permitir encontrar facilmente as referências dos requisitos, de modo a facilitar a compreensão e identificação da origem ou bases para definição dos requisitos.

As partes do documento contendo uma especificação de requisitos de software são indicadas pela norma IEEE Std 830-1998, conforme apresentado a seguir.

Conteúdo
1. Introdução
1.1 Objetivo
1.2 Escopo
1.3 Definições, acrônimos e abreviações
1.4 Referências
1.5 Visão Geral
2. Descrição Geral
2.1 Perspectiva do produto
2.2 Funções do produto
2.3 Características dos usuários
2.4 Limitações
2.5 Suposições e dependências
3. Requisitos específicos
Apêndices
Índice

**Figura 15 – Protótipo de um modelo de especificação de requisitos (IEEE, 1998)**

## **Capítulo 3**

### **Análise de Ambientes Baseados em Ontologias**

### **3 Análise de Ambientes de Autoria Baseados em Ontologias**

Neste capítulo analisaremos os ambientes de autoria existentes que fazem referência à Web Semântica e que são baseados em ontologias. Uma das etapas do processo de análise é a definição de um modelo de análise, o qual foi elaborado a partir de critérios, padrões, recomendações e requisitos existentes, utilizados para análise e avaliação de ambientes. Os critérios foram buscados na literatura, listados no Anexo I desta tese, e organizados em quatro categorias, apresentadas neste capítulo:

- ambientes de autoria;
- ontologias;
- W3C; e
- software.

A partir do estudo desses critérios, selecionamos um conjunto de características de cada categoria e formamos um modelo de análise de ambientes de autoria baseados em ontologias para a Web Semântica. O resultado desta análise permitiu a identificação de requisitos e também revelou tendências e habilidades que usuários desses ambientes devem ter para deles usufruir.

Com esse trabalho, contribuímos para a identificação de atividades inerentes à produção de documentos semanticamente representados, desenvolvemos um critério de análise de ambientes de autoria e com ele identificar características de sistemas existentes, assim como o delineamos orientações para o desenvolvimento de novos sistemas. Para tanto, identificamos uma série de requisitos que podem ser adotados por desenvolvedores de ambientes de autoria e contribuir para que a produção de conhecimento na Web Semântica seja feita de forma estruturada e relacionada a ontologias no momento da autoria do documento.

A metodologia usada para alcançar os objetivos nesta pesquisa consiste na busca por métodos de análise e avaliação nos campos da engenharia de software, aquisição de conhecimento e ferramentas de autoria. Neste capítulo iremos apresentar, analisar e extrair características que poderão ser usadas para as propostas de soluções aqui apresentadas. Como resultado, propomos um critério específico para analisar ambientes de autoria baseados em ontologias para a Web Semântica.

Os critérios de análise tornam-se cada vez mais necessários à medida que novos sistemas são desenvolvidos. Alguns padrões já existentes e podem dar apoio à formação de um olhar crítico que identifique novas tecnologias baseadas em abordagens específicas. Contudo, definir tal critério não é uma tarefa simples e esforços significativos devem ser empregados para juntar estes padrões e propor novos modelos, a fim de auxiliar os trabalhos de análise desses ambientes.

### **3.1 Critérios de análise dos ambientes de autoria**

Algumas pesquisas têm orientado a tarefa de avaliação de tecnologias de autoria (BOSTOCK, 1994). CRISTEA (2004) conduziu uma avaliação de um sistema de autoria de uma hipermídia adaptativa, por meio do teste de sistemas em conjunto com os autores para ver, na prática, como a eficiência da autoria tem sido afetada em termos das específicas funcionalidades do sistema. Primeiramente foram levantadas algumas características dos usuários do sistema e definidos objetivos da avaliação com base nessas características (CRISTEA, 2003). Os métodos de avaliação baseiam-se em questionários criados por usuários para serem respondidos durante o uso do sistema. Os aspectos observados nos pontos de vistas buscados são: colaboração, completude, adaptabilidade, e variação do design; e com relação à interface do usuário foram buscados os seguintes: facilidade de uso, robustez e complexidade das interfaces. Os experimentos revelaram descobertas sobre os próximos passos a serem tomados no processo de implementação de modo a preencher as características definidas e orientações para a melhoria do processo de implementação.

SOBIESIAK (1996) apresenta questões-chave que podem ser consideradas quando do desenvolvimento de ferramentas de autoria para autores que não são especialistas, ou não são profissionais, ou para aqueles que desempenham essa função eventualmente. MACKNIGHT et al (1989) criaram uma ferramenta para a avaliação dos sistemas de autoria de materiais educacionais. Eles consideraram que a avaliação de sistemas de autoria deve prever fatores como facilidade de uso, transportabilidade, redução de gastos e tempo envolvido na criação de materiais educativos. O modelo de avaliação consiste em *“procedimentos e um conjunto comum de tarefas para fatores de desempenho e capacidades de processamento”*. As tarefas são efeitos instrucionais criados pelos autores via interação com o

sistema. Elas são medidas sob três dimensões: i) funcionalidade, ii) flexibilidade e iii) produtividade.

A funcionalidade é medida pela variação de funções apoiadas pelo sistema. Os autores dividiram algumas áreas funcionais e apresentaram conjuntos de tarefas envolvendo em cada uma delas quatro dos exemplos abaixo citados: i) ambiente de autoria; criação de texto; edição de texto; criação gráfica; edição; animação; vídeo interativo; estratégia institucional; cálculos; administração de alunos. Os sistemas poderiam assim ser pontuados de acordo com parâmetros existentes e usando uma “razão de funcionalidade”. Para calcular a funcionalidade, por exemplo, divide-se a quantidade de tarefas implementadas pelas quantidades existentes de tarefas em uma área funcional fornecida pela ferramenta de avaliação. A flexibilidade é calculada pelo número de tarefas implementadas pelo sistema, dividido pelo número total de passos numa função específica. A produtividade é medida pelo grau de facilidade do uso do sistema. Baseada na interface do usuário, no número de passos para implementar uma tarefa e no menor número de passos para criar um objeto específico (MACKNIGHT, 1989).

O modelo de análise proposto nesta tese está voltado a identificar formas de interação entre o usuário e o sistema com relação ao aspecto do uso do sistema para criar o conteúdo, relacionar esse conteúdo com outras fontes, ontologias e criar metadados, anotações etc. Está focado nos aspectos sintáticos e semânticos, ou seja, na maneira em que a informação é apresentada pelo sistema ao usuário.

### ***3.2 Critérios para análise de tecnologias baseadas em ontologia***

Nesta seção apresentaremos algumas abordagens e iniciativas voltadas para análise, avaliação e comparação de tecnologias baseadas em ontologias. Apresentamos os critérios desenvolvidos pelos grupos de pesquisa aqui citados, para facilitar a organização e identificação de características comuns entre eles, o que contribuiu para construir a proposta de critérios utilizada nessa análise e para a modelagem de um novo ambiente de autoria para a Web Semântica.

Os ambientes de autoria baseados em ontologia são tecnologias emergentes utilizadas para desenvolver conteúdos para a Web Semântica com a ajuda de ontologias e metadados. Para que possam ser processados pela Web Semântica e fazer uso dos seus potenciais de representação, os produtos de informação gerados por essas ferramentas precisam estar de acordo com as recomendações de

representação do conhecimento propostas pelo W3C, tais como: formatos de arquivo XML, geração de metadados, acessibilidade, entre outros. Outras recomendações oriundas da análise de sistemas também podem contribuir para o desenvolvimento de critérios de análise. Nesta seção apresentaremos algumas abordagens específicas voltadas aos aspectos de análise, comparação e avaliação de tecnologias baseadas em ontologias.

A avaliação de tecnologias de ontologias começou a ser realizada quando a primeiras aplicações de ontologias apareceram. Começou com GRUBBER (1994), que descreveu cinco critérios gerais de design sobre como as ontologias deveriam ser desenvolvidas em termos de engenharia de design para dar apoio ao compartilhamento de conhecimento. Os trabalhos conduzidos por GÓMEZ-PÉREZ (1995) têm contribuído para o desenvolvimento dessa tarefa específica, revelando que é importante considerar a avaliação de ontologias como processo crítico na integração dessa tecnologia em sua real aplicação. A autora propõe um primeiro critério para verificar a tecnologia de compartilhamento de conhecimento (*knowledge sharing*), para verificar que a definição esteja correta na ontologia, e verificar se os ambientes de software e de documentação relativos são considerados durante o ciclo de vida das ontologias. Basicamente as ontologias são verificadas em relação à estrutura, léxico, sintaxe e conteúdo. O critério para cada nível respectivo é adequação, correção, consistência, completude, concisão, expansão e sensibilidade.

GÓMEZ-PÉREZ (1995) justifica seu trabalho de avaliação de ontologias, dizendo que *“a avaliação de ontologias assim como a avaliação de documentos e ambientes de software é crítica para a integração dessa tecnologia em aplicações reais”*. *“Se as definições da ontologia e o ambientes de software não forem suficientemente avaliados, a comunicação entre agentes de softwares talvez não alcance sucesso”*. Ela aponta sobre a consistência da arquitetura da ontologia; a correção do léxico e da sintaxe das definições; consistência, completeza, concisão, expansão e “sensibilidade” do conteúdo das ontologias e suas definições. No que se refere ao conteúdo, a autora apresenta a necessidade de desenhar: uma definição formal interna, metafísica e de consistência inferida; e uma informal definição da completude e estereótipos de erros que fazem relações, funções e classes incompletas; uma definição informal de concisão, e tipos de erros que fazem a

ontologia tornar-se redundante; e que tipo de verificação tem de ser tomada quando decisões são adicionadas ou modificadas em uma ontologia.

Um estudo de *benchmarking* com tecnologias de ontologias foi realizado recentemente pelo projeto ONTOWEB (2002), no qual os autores definiram categorias de ferramentas baseadas em ontologias, a grande maioria relacionada à engenharia de ontologias, ou seja, à atividade de construir e gerenciar as ontologias. Com base nessas características, definiram um conjunto específico de critérios de avaliação de ferramentas de ontologia (ONTOWEB, 2004). As ferramentas de ontologia foram divididas nas seguintes categorias: ferramentas de construção, união, integração e avaliação de ontologias; anotação baseada em ontologias; e armazenamento e *query* de ontologias. Como resultado, o projeto também apresentou um critério específico de avaliação de ferramentas de construção de ontologias – incluindo a maioria das ferramentas de edição de ontologias conhecidas no mercado, como *Protégé*, *WebODE*, *OilEd*, *KAON*, *Ontolingua*, etc. (ONTOWEB, 2004).

As aplicações de ontologias podem ser também classificadas e organizadas em categorias de grupos específicos, dependendo do objetivo específico da aplicação, como por exemplo o modelo oferecido por USCHOLD (1999): i) autoria neutra, ii) acesso comum à informação e iii) indexação para pesquisa. Esse modelo também indica as tecnologias de apoio e o papel dos atores que participam do sistema que está sendo utilizado.

ANGELE (2003) propõe um *framework* de avaliação de ontologias e de tecnologias relacionadas, divididas em dois focos principais: propriedades de ontologias e propriedades de tecnologias. As propriedades de ontologias estão relacionadas à conformidade da linguagem (sintaxe) e à consistência (semântica) enquanto que as propriedades de tecnologias estão relacionadas a interoperabilidade (semântica), movimentação da ontologia, performance, alocação de memória, escalabilidade, integração em módulos, conectores e interfaces.

No modelo proposto por ANGELE (2003), consideraremos para o critério de análise, duas categorias de ferramentas: construção de ontologias e ferramentas de anotação baseadas em ontologias. As demais categorias não serão consideradas pois estão fora do escopo dessa análise: união de ontologias e ferramentas de integração, ferramentas de avaliação de ontologias, armazenamento de ontologias e ferramentas para elaboração de *query*. Esses aspectos serão observados de

maneira geral, por exemplo, identificaremos se as ferramentas analisadas apresentam essas características.

DENNY (2004) conduziu uma pesquisa de ferramentas de edição de ontologias, definindo critérios para a coleta de informação a partir das ferramentas, e, como resultado, apresentou um conjunto de características que podem melhorar o desempenho das ferramentas de edição de ontologias.

GRUBBER (1993) introduz princípios para o design de ontologias usadas para compartilhamento de conhecimento. SURE (2002) apresentou uma metodologia para o desenvolvimento e uso de aplicações de gestão de conhecimento baseadas em ontologias. TALLIS (1999) apresentou estudos de usuários de ferramentas de aquisição de conhecimento, por meio do relato de metodologias e lições aprendidas. CORCHO (2003) apresentou um experimento de avaliação de ferramentas de interoperabilidade em ontologia com o *workbench* de engenharia em ontologia – WebODE.

Por fim, indicamos nesta tese como esses padrões acima citados podem ser agrupados para criar um conjunto específico de critérios que podem ser usados para avaliar ambientes de autoria baseados em ontologia. O resultado gerou a proposta de um novo padrão a ser usado na avaliação desses ambientes de autoria, de modo a dar suporte aos produtores de novos ambientes.

Concluimos com essas abordagens que a maioria dos critérios propostos ou de avaliação de *frameworks* estão direcionados para a tarefa de pontuação (*scoring*) da eficiência das ferramentas de construção e gerenciamento de ontologias. Não existem abordagens que se relacionem com a análise de sistemas que possuem ontologias como elemento intermediário para se alcançarem outras tarefas. Para compensar a ausência desse tipo de abordagem, estamos propondo um critério de análise de ambientes de autoria baseados em ontologia, apresentado no item 3.5 deste capítulo.

### **3.3 Critérios para análise de conformidade com recomendações do W3C**

Com relação aos resultados gerados pelos ambientes de autoria, os documentos produzidos na Web Semântica devem estar de acordo com as recomendações emitidas pelo W3C. O Grupo de Trabalho HTML (W3C, 2006) produziu um conjunto de especificações, discutidas e acordadas pelos seus membros constituintes, que orientam os autores a partir de um conjunto de

características que os documentos em html e xml devem ter para serem processados pela Web Semântica. Esse grupo de trabalho apresenta formas de garantir a confiança e fidelidade de documentos com padrões da Web Semântica, por meio de um serviço de validação de marcação (W3C, 2006), que consta de uma aplicação que automaticamente checa se as *tags* dos documentos em html ou xhtml estão de acordo com as recomendações, garantindo que os editores de html são interoperáveis com as outras aplicações da Web e também checa se os *outputs* estão de acordo com as recomendações. Também é interessante mencionar um aplicativo chamado HTML Tidy (W3C, 1999), um software standalone que faz tradução automática de documentos em HTML para documentos em formato XHTML. Ele verifica também a edição das marcações dos documentos HTML, consertando erros de *tags* e demonstrando ao autor onde estão os erros para correções manuais ou automáticas, caso seja possível e necessário. A ferramenta também faz a verificação dos requisitos de acessibilidade, ou seja, se os documentos produzidos estão de acordo com as recomendações do grupo de trabalho de acessibilidade para a Web Semântica – WAI (W3C, 2005), responsável por definir os requisitos para que as ferramentas sejam utilizadas por usuários portadores de necessidades especiais. O que aprendemos com o HTML Tidy é que com ele é possível checar a consistência de documentos, comparando as estruturas com formatos-padrão estabelecidos. Apesar de ser customizado pelo autor, HTML Tidy é restrito a html e tem várias limitações ao lidar com documentos em XML, pois não reconhece todas as características desse formato.

Com relação à especificação de linguagem de edição, de acordo com W3C (2005) os ambientes de autoria para a Web Semântica devem explicitar qual idioma primário está sendo utilizado no conteúdo do documento, assim como o idioma utilizado no processamento do texto. A codificação da linguagem não significa apenas codificação de caracteres ou direção do texto, logo, o idioma do documento e a linguagem utilizada para a sua redação devem ser explicitados na estrutura de metadados. Essa recomendação do W3C também oferece algumas instruções sobre como informar a linguagem de processamento de texto, e também dizer se o documento é multilíngüe.

Na Web Semântica, as ontologias são escritas por meio de linguagens especificamente desenvolvidas para representar o conhecimento por meio de relações lógicas. A linguagem OWL (W3C, 2004) é hoje a linguagem padrão para a

representação do conhecimento na Web Semântica. Foi criada para ser utilizada em um ambiente onde as informações contidas em documentos precisam ser processadas pelos sistemas, em oposição a situações onde o conteúdo é apenas apresentado a humanos. A linguagem OWL pode ser usada para representar explicitamente o significado de termos em vocabulários e as relações entre esses termos, por meio de ontologias. A linguagem OWL é uma evolução da DAML+OIL, e suas recomendações usam tipos simples dos esquemas de XML (W3C, 2005).

Há também recomendações do W3C (2005) para documentos compostos, que são documentos que contêm diferentes linguagens de representação para referência a outros documentos, como por exemplo um documento em .html um gráfico em .xls e uma imagem .gif. Essa especificação orienta ambientes de produção de conteúdo e ferramentas de autoria, agentes, navegadores e *plug-ins* etc. Os documentos compostos podem ser produzidos por uma variedade de meios como, por exemplo: editor de texto comum, geração dinâmica de documentos, linguagens de componentes ou as ferramentas de autoria de documentos compostos. Seguem alguns exemplos (W3C, 2005).

- Ferramentas de autoria centrada em documentos – para a criação de conteúdo que combina texto estático e gráficos para fins de publicação na Web.
- Ferramentas de autoria centrada em multimídia – para a criação de conteúdo interativo (como um filme ou jogo, por exemplo).
- Ferramentas de autoria centrada em aplicações – para a criação de interfaces de usuários.
- Ferramentas de autoria centrada em formulários – para a criação de formulários-padrão (modelos) de dados.
- Ferramentas de autoria independentes de dispositivos – para a criação de conteúdo que pode ser adaptado a diferentes requisitos de usuários e que interagem com diferentes tipos de mídias.

O W3C oferece também diretrizes para o desenvolvimento de ferramentas de autoria de conteúdo para a Web que sejam mais acessíveis para pessoas com necessidades especiais (W3C, 2005). Uma ferramenta de autoria que estiver em conformidade com essas diretrizes pode promover a acessibilidade ao oferecer uma interface de autoria acessível a autores com deficiência, assim como capacitar, dar suporte e promover a produção de conteúdo acessível para os usuários da Web. É

importante notar que muitas ferramentas de autoria já apresentam algum tipo de função de acessibilidade compatível com algumas das características exigidas, como por exemplo, alguns editores de HTML têm tanto funções de autoria WYSIWYG (*what you see is what you get*) e também funções de autoria orientadas a objeto, como o Word da Microsoft.

O W3C tem buscado definir padrões para a Web Semântica por meio da publicação de recomendações desenvolvidas por grupos de trabalho específicos para tarefas específicas tais como: representação do conhecimento, acessibilidade, computação móvel, cooperação de agentes, etc. Ainda, devido à importância do momento de autoria, quando milhões de páginas são produzidas diariamente por todo o mundo, sugerimos o desenvolvimento de orientações e recomendações para facilitar essa tarefa e ajudar a gerar documentos que possam ser processados pela Web Semântica. Esse tipo de iniciativa mostra a importância de seu aspecto de “inclusão social”, assim como a necessidade por orientação de projetos que possam ser usados por desenvolvedores dessas ferramentas. Isso torna a estrutura e o conteúdo mais fáceis de serem desenvolvidos uma vez que podem contar com orientações didáticas acordadas por grupos multidisciplinares.

### **3.4 Critérios para análise de ambientes de software**

A norma ISO 14598-1(ISO,1998) propõe um conjunto de padrões para a avaliação de softwares e que em geral pode ser usada para a avaliação de sistemas de processamento de linguagem natural assim como para muitos outros tipos. Esse padrão estabelece que no processo de avaliação do sistema existem, basicamente, quatro estágios a serem considerados quando da condução dos estudos (King, 2003). O primeiro estágio é sobre os requisitos de avaliação: estabelece a proposta de avaliação; identifica os tipos de produtos a serem avaliados; e especifica o modelo de qualidade. O segundo estágio é sobre a especificidade da avaliação em termos métricos, pontuação e critério para análise. O terceiro é sobre o desenho da avaliação em termos de cronograma. O quarto e último estágio é sobre a execução da avaliação, dividida em fases de medida, pontuação e análise. Esses quatro estágios são construídos a partir de um modelo de qualidade, o qual é um conjunto de qualidades características que a ISO 9126 (ISO, 2001 e SCALET, 2000) propõe para o processo de avaliação: funcionalidade, confiabilidade, usabilidade, eficiência, manutenção e portabilidade. KING (2003) também disserta sobre modelos de

avaliação de softwares de processamento de linguagem natural, baseados nos padrões da ISO.

Atualmente as tarefas básicas de definição de requisitos para a projeção e desenvolvimento de sistemas podem ser guiadas por diversas abordagens metodológicas (SAWYER 2001, DAVIS 1993, IEEE, 1988). O trabalho de análise de requisitos na engenharia de software é fundamental para se ter um entendimento dos componentes de um sistema e as relações entre eles. Diversas metodologias abordam o assunto (KITCHENHAM, 1999) e esse entendimento é atingido por meio da experimentação, ou seja, da realização de testes de funcionalidade. A experimentação é feita para ajudar na avaliação dos sistemas, de modo a prever, entender, controlar e melhorar o processo de desenvolvimento de softwares e seus produtos (BASILI, 1986). Com a avaliação, modelos do processo funcional dos sistemas e seus produtos são construídos; hipóteses são levantadas e testadas, e o resultado é utilizado para refinar essas hipóteses ou gerar novas outras.

Para NORMAN (2000, p.1):

*“as técnicas de avaliação tradicionais, freqüentemente guiadas por modelos baseados em regressão, para estimativa de custos e previsão de defeitos, provêm pouco apoio para os gerentes que desejam utilizar tais medidas para analisar e diminuir riscos. O futuro das técnicas de métrica e avaliação de sistemas está no uso relativamente simples de métricas existentes para construir ferramentas de gerenciamento para gestão do apoio à tomada de decisão, combinando diferentes aspectos do desenvolvimento e teste de softwares e dar condições aos gerentes para fazer diferentes tipos de previsões, análises e trocas durante o ciclo de vida do sistema.”*

Neste capítulo, os esforços na definição dos critérios de avaliação que podem ser utilizados para análise de ambientes de autoria baseados em ontologia foram apresentados. Até o presente momento, alguns grupos de sistemas de engenharia desenvolveram critérios de avaliação gerais. Especialistas em engenharia de ontologia têm também desenvolvido meios para pontuar as ferramentas de ontologia, e que esses critérios propostos ou modelos de avaliação têm como objetivo pontuar a eficiência das ferramentas de construção e gerenciamento de ontologias, exclusivamente. Apresentando essas iniciativas relacionadas propomos desenvolver um critério específico que poderá ajudar na tarefa de identificar e

pontuar sistemas de autoria em Web Semântica. Com os resultados, seremos capazes de identificar o estado da arte na autoria.

Nesse trabalho buscamos preencher essa lacuna, apresentando uma proposta de critérios de análise que possa contribuir para a seleção ou desenvolvimento de ambientes de autoria baseados em ontologias. Os resultados dessa pesquisa foram compilados e sistematizados, por meio da Tabela 1, contida no anexo I da tese. A lista em anexo foi mantida em língua inglesa para facilitar a identificação dos termos originais dos critérios que foram adotados nas análises apresentadas, os quais foram traduzidos em língua portuguesa pela primeira vez nesse trabalho.

A partir da análise dessas abordagens, apresentamos até então uma base conceitual e empírica para estabelecimento do critério de análise proposto, o qual permitirá a análise dos ambientes de autoria baseados em ontologia para a Web Semântica. Vale ressaltar que esse processo irá também servir para validar os próprios critérios de análise que estamos propondo na presente pesquisa e que formarão a base para a construção do modelo de produção de documentos baseado em ontologias para a Web Semântica.

### ***3.5 Proposta de critério de análise de ferramentas de autoria baseadas em ontologias***

Para a presente análise, as funcionalidades dos sistemas foram classificadas em termos de características existentes (dimensão do sistema) estrutura do componente base (interoperabilidade com outros sistemas), dentre outras. Nessa análise, identificamos tarefas comuns nos sistemas e descrever como elas são realizadas em termos de funcionalidade. As tarefas identificadas nos sistemas são aquelas relacionadas à composição do texto com o uso de ontologias.

Para completar o critério de análise que buscamos propor, adaptamos também alguns modelos de avaliação gerais relativos a ambientes de autoria (CRISTEA, 2004; MCKNIGHT, 1989; HANDSCHUH, 2002).

A partir do estudo das abordagens apresentadas nas seções anteriores e compiladas no Anexo I da tese, foi selecionado um conjunto de características identificadas nos sistemas em análise, para fins de definição de um modelo de análise específico para a avaliação de ferramentas de autoria baseada em ontologias, proposto a seguir.

**Tabela 1 – Modelo de análise de requisitos**

<b>Características Identificadas</b>
<b>Ambiente de Autoria</b> Editor de texto Gerenciamento de documentação Suporte a publicação Extração de informação Geração de conteúdo Suporte a mais de um usuário Suporte léxico Criação de gráficos Suporte para formatação de documento Restrição de entrada de dados Interação com usuário Ambiente de pesquisa e descoberta Conceitos interligados Navegação em fontes externas Documento em html Relacionado a outras tecnologias ou formato de arquivos
<b>Gerenciamento da Ontologia</b> Guia e visualização da ontologia Meta-ontologia Crawler Serviço de anotação automática Inferências Movimentação da ontologia Integração em módulos Conexões e interface - relações semânticas Múltiplas ontologias União de ontologias Alinhamento de ontologias Controle de versão Suporte para contextos específicos Raciocínio e solução de problema Usuário popula/edita ontologia
<b>Conformidade com W3C</b> Ontologias em OWL Docs baseados em XML Geração de metadado Reconhecimento de metadados Especifica idioma do contexto Especifica linguagem de processamento XML Schema em RDF e OWL Acessibilidade Documento composto
<b>Questões de Software</b> Interoperabilidade Documentação de suporte a usuários Adequabilidade Segurança Simple interface com usuário Interface orientada a objeto Fácil de usar Fácil de aprender Atratividade

As informações foram coletadas a partir do uso dos sistemas e da análise dos documentos referentes, como manuais, tutoriais, artigos científicos, telas de sistemas, assim como da análise de informações providas pelos desenvolvedores das ferramentas. As características selecionadas foram agrupadas em 4 grandes categorias: autoria, gerenciamento da ontologia, conformidade com o W3C e questões gerais de software.

Foram identificadas características relacionadas a três aspectos específicos: se a característica necessária estava ou não implementada, se estava implementada parcialmente ou se não estava implementada. O principal objetivo é identificar quais são as características presentes nas ferramentas, quais características ainda não foram implementadas, considerando as últimas recomendações providas, o estado da arte assim como novas possibilidades e inovações. O objetivo dessa análise não foi o de pontuar ou julgar as ferramentas em um ranking de características, mas sim de identificar a presença ou ausência das características.

Das abordagens práticas encontradas, listadas no Anexo I, foram selecionados, analisados e adaptados os critérios disponíveis para a realização dessa pesquisa. As abordagens que tratam especificamente sobre engenharia de ontologias foram consideradas parcialmente, uma vez que o objetivo da pesquisa não é avaliar a qualidade da engenharia realizada. Com relação às questões de engenharia de software, foram estudadas abordagens de métrica de software e consolidados os critérios utilizados por BASILI (1986) e NORMAN (2000), assim como nos padrões da ISO/IEC (2001) a qual aborda sobre critérios de avaliação de qualidade na engenharia de software, tais como usabilidade, facilidade de uso, interoperabilidade, segurança, documentação e tutoriais, adequação aos propósitos, complexidade da interface e habilidades dos usuários, dentre outras.

### **3.6 Aplicação do modelo nos sistemas analisados**

No site *Semantic Annotation and Authoring* do W3C (Semantic, 2005), há uma listagem de 13 ferramentas, das quais 10 são utilizadas exclusivamente para anotação semântica e apenas três delas apresentam características e funções de autoria. São elas: Semantic Word (TALLIS, 2003), CREAM Ont-o-Mat (HANDSCHUH, 2002) e Smore (KALYANPUR, 2005). Em uma pesquisa mais ampla (OLIVEIRA, 2006), descobrimos três outras ferramentas que também

apresentam características de autoria Cohse (CARR, 2001), WickOffice (CARR, 2004) e Trellis (GIL, 2002). Apresentamos abaixo uma descrição individual de cada uma delas.

O WickOffice (CARR, 2004) é um ambiente de produção de conhecimento baseado em um dado conjunto de ontologias que dão suporte aos usuários no preenchimento de formulários de solicitação de financiamento de pesquisas. O sistema busca informações para compor documentos baseados em contextos específicos. Os documentos podem ser escritos por meio de um serviço de auxílio ao autor, onde eles são guiados a selecionar as classes e instâncias corretas das ontologias e estruturar os documentos de acordo com uma estrutura e critérios restritos. É um ambiente fechado de autoria, o qual impõe aos usuários as possíveis respostas dos formulários, sugerindo a resposta adequada a ser dada.

O Semantic Word (TALLIS, 2003) é uma ferramenta de anotação semântica que auxilia os autores a anotar conteúdos durante a autoria. O sistema é flexível ao permitir que os usuários livremente adicionem, nas ontologias, novas instâncias ou atribua novas relações entre elas. O sistema também oferece opção de autoria baseada em formulários ou modelos. A diferença entre Semantic Word e o anterior, WickOffice, é com relação à opção de população da ontologia, que, no último, é livre, enquanto que no primeiro são impostas condições e limitações ao autor no uso das instâncias e relações criadas na ontologia.

O sistema Smore (KALYANPUR, 2005) permite aos autores criar simples páginas Web, durante a marcação de documentos. Os autores necessitam de um conhecimento mínimo de sintaxe e familiaridade com a linguagem OWL. O sistema permite a criação de classes, propriedades e indivíduos a partir de ontologias existentes, realizar edição ontológica básica, ou mesmo criar uma nova ontologia a partir dos termos dos documentos da Web.

O Cream Ont-o-mat (HANDSCHUH, 2002) é um ambiente para anotação e autoria na Web Semântica, que oferece a integração entre anotação e autoria, durante a navegação na Web. A autoria de páginas é limitada a um ambiente específico de edição de html, sem as opções de edição de textos existentes hoje em editores de textos.

O sistema Cohse (CARR, 2001), apresentado no capítulo anterior, é um protótipo de um sistema Web que integra uma ontologia de domínio específica, construída a partir de descrições lógicas que dão suporte à classificação de termos,

gerenciamento automático da terminologia utilizada e também catalogação do documento para dar suporte à autoria e à navegação semântica. Apresenta um serviço de raciocínio ontológico com o qual ele representa modelos conceituais de termos e seus relacionamentos. Apresenta também um serviço de *link* de hipermídia aberta, baseado em documentos da Web, que oferece representação dos documentos por meio de metadados que descrevem seus conteúdos.

O Trellis (GIL, 2002) é um sistema que permite aos usuários adicionar comentários sobre documentos analisados, em um ambiente interativo que oferece opções de redação de opiniões e decisões analíticas sobre documentos e fontes de informação. Os usuários podem fazer conexões semânticas entre partes de um documento e entre outros documentos, criando-se assim uma rede de comentários sobre esses documentos e tornando-os disponíveis para outros usuários e revisores. Outros sistemas, como o SEAS (LAWRENCE, 2001) e o Annotea (KOIVUNEN, 2001) apresentam abordagens similares à do Trellis e são também complementares às tarefas de anotação e autoria. SEAS é mais específico para domínios enquanto que o Trellis é mais genérico e é mais voltado para a interpretação de documentos para tomada de decisão, permitindo a autoria de comentários. O Annotea é uma infra-estrutura de anotação semântica para os usuários adicionarem comentários sobre as fontes de informação, produzir resumos ou gerenciar a cadeia de comentários feitos por outros usuários. O sistema Claimaker (UREN, 2003) também apresenta essas mesmas funcionalidades, provendo uma estrutura de modelagem de argumentos sobre documentos existentes, por meio de técnicas de engenharia de conhecimento e visualização das redes de argumentação. Essas redes são criadas por meio da anotação feita nos textos, a partir de conceitos reconhecidos nas ontologias. Permite a criação de uma ontologia de discurso, por parte dos autores, para trabalhos colaborativos de revisão de pares.

A partir desse conjunto de ferramentas, selecionamos seis delas que representam funcionalidades específicas, dentro do escopo de autoria e anotação semântica, para fins de identificação de características, com base no modelo de critério definido neste capítulo, e apresentado na Tabela 2, a seguir: WickOffice, SemanticWord, Smore, CreamOnt-o-mat, Cohse e Trellis.

Tabela 2: Consolidação da Análise de Ferramentas

Ferramentas	Wick Office	Semantic Word	Smore	Cream Ont-o-mat	Cohse	Trellis
<b>Características</b>						
<b>Ambiente de Autoria</b>						
Editor de texto	☺	☺	☺	☺	☺	☺
Gerenciamento de documentação	☺	■	■	☺	☺	☺
Suporte a publicação	☺	☺	☺	☺	☺	☺
Extração de informação	☺	☺	☺	☺	☺	■
Geração de conteúdo	☺	☺	☺	☺	☺	■
Suporte a mais de um usuário	■	■	■	■	●	●
Suporte léxico	☺	☺	☺	☺	☺	■
Criação de gráficos	■	■	●	■	■	■
Suporte para formatação de documento	☺	☺	●	☺	●	■
Restrição de entrada de dados	☺	●	■	☺	●	■
Interação com usuário	●	☺	☺	●	☺	☺
Ambiente de pesquisa e descoberta	●	☺	☺	☺	☺	☺
Conceitos interligados	☺	☺	☺	●	☺	●
Navegação em fontes externas	■	☺	☺	☺	☺	☺
Documento em html	■	■	☺	☺	☺	☺
Relacionado a outras tecnologias ou formato de arquivos	☺	☺	☺	☺	■	☺
<b>Gerenciamento da Ontologia</b>						
Guia e visualização da ontologia	■	☺	☺	☺	☺	●
Meta-ontologia	■	☺	☺	☺	☺	●
Crawler	☺	☺	☺	☺	☺	■
Serviço de anotação automática	☺	☺	☺	●	☺	■
Inferências	■	■	☺	☺	☺	■
Movimentação da ontologia	■	☺	☺	●	☺	☺
Integração em módulos	☺	☺	☺	☺	☺	☺
Conexões e interface - relações semânticas	■	☺	☺	☺	☺	☺
Múltiplas ontologias	☺	☺	☺	■	☺	☺
União de ontologias	■	■	■	■	■	■
Alinhamento de ontologias	■	●	●	●	■	■
Controle de versão	■	■	■	■	■	■
Suporte para contextos específicos	☺	■	☺	☺	☺	☺
Raciocínio e solução de problema	■	■	■	■	☺	■
Usuário popula/edita ontologia	☺	☺	☺	☺	☺	☺
<b>Conformidade com W3C</b>						
Ontologias em OWL	■	■	☺	■	■	☺
Docs baseados em XML	☺	☺	☺	☺	☺	☺
Geração de metadados	☺	☺	☺	☺	☺	☺
Reconhecimento de metadados	☺	☺	☺	☺	☺	☺
Especifica idioma do contexto	■	■	■	■	■	■
Especifica linguagem de processamento	☺	☺	☺	☺	■	☺
XML Schema em RDF e OWL	●	●	☺	●	■	☺
Acessibilidade	●	●	●	●	●	●
Documento composto	■	■	■	■	■	■
<b>Questões de Software</b>						
Interoperabilidade	☺	☺	☺	●	■	■
Documentação de suporte a usuários	●	●	■	☺	●	●
Adequabilidade	☺	☺	☺	☺	●	☺
Segurança	■	■	■	■	■	●
Simple interface com usuário	☺	●	●	●	☺	■
Interface orientada a objeto	☺	☺	☺	●	☺	●
Fácil de usar	●	■	■	■	●	●
Fácil de aprender	☺	☺	●	●	●	●
Atratividade	☺	☺	☺	■	☺	■

Legenda:  
 ☺ Característica implementada  
 ● Característica parcialmente implementada  
 ■ Característica ainda não implementada

### **3.7 Resultados encontrados**

A partir da análise dos resultados, identificamos que nenhuma das ferramentas apresenta a solução como um todo. Do ponto de vista da formatação do documento, os autores são limitados a utilizar formatos .doc ou .html. Outros formatos de edição ainda não estão disponíveis, o que restringe os ambientes e suas características de edição disponíveis. Aos usuários resta ainda ter de integrar ambientes de modo a atingir um bom nível de redação e edição de documentos em ambientes de autoria, uma vez que as ferramentas não vislumbram o aspecto da interoperabilidade.

As ferramentas são mais voltadas para anotação semântica do que para autoria baseada em ontologia. A anotação é um dos passos preliminares para se fazer a representação da informação com uso de ontologias. Talvez seja esse o motivo pela escassez de tecnologias de autoria até o momento, ficando a autoria em um segundo plano, exceto na proposta da ferramenta Semantic Word e WickOffice, as quais integram anotação com o ambiente Microsoft Word.

A extração da informação é uma atividade da engenharia lingüística essencial para o reconhecimento dos conceitos e seus termos, para a conseqüente criação de suas relações, que formam as ontologias. No entanto, pode tornar-se uma tarefa deveras complexa, caso a extração não seja feita automaticamente e via um modelo integrado. As técnicas precisas de extração da informação podem ser muito complexas e podem também sobrecarregar os autores caso isso seja designado para eles fazerem. Logo, deve ser deixado para os autores o mínimo de extração possível, de preferência apenas a validação ou a correção de inconsistências semânticas e sintáticas.

O suporte a multi-usuários é uma característica importante em ambientes de produção colaborativa de documentos. A produção de conhecimento na atual Sociedade da Informação é uma atividade que envolve acesso à informação, análise, discussão, produção, publicação etc., e geralmente isso é feito por meio de duas ou mais pessoas, quando não uma pessoa ou mais para cada etapa, logo, recomenda-se que os ambientes em desenvolvimento considerem o uso simultâneo de mais de um usuário no sistema.

A informação pode ser produzida livremente ou dentro de padrões específicos, dependendo do contexto. Ambientes abertos e flexíveis podem deixar a

opção para o autor em definir o tipo de representação a ser realizada, que termo melhor identifica um conceito, as relações entre as instâncias de uma ontologia etc. No entanto, quando se fala de ambientes fechados, onde regras rígidas se aplicam no controle de informações contidas em documentos, deve-se levar em conta o desenvolvimento de características que possam restringir algumas ações, como por exemplo, o uso de termos genéricos ou preferidos, ou evitar o uso de outros, por meio de tesauros ou equivalentes. A partir das regras de cada contexto, o desenvolvedor deve poder considerar o grau de restrição que as opções de uso do sistema podem ter. Como exemplo, citamos o que ocorre com o sistema WickOffice, o qual oferece formulários-padrão (modelos ou *templates*) para direcionar o usuário no tipo de formulário e estrutura do documento a ser gerado, assim como sugere o uso da terminologia correta e em uso corrente para representar o conteúdo equivalente. É uma característica particular, em comparação aos sistemas Trellis e Smore, os quais são sistemas que deixam em aberto para o autor definir que tipo de relações podem ser estabelecidas entre conceitos para propósitos particulares.

Finalmente, também notamos que os sistemas não apresentam todas as características básicas recomendadas pelo W3C (2006), assim como aspectos de engenharia de software e gerenciamento de ontologias. É necessário discutir essas questões de modo a prover facilidades de uso e aprendizado dessas ferramentas para os usuários, aumentar o nível de operabilidade e usabilidade entre elas e outras de propósito comum ou complementar, ou seja, a autoria baseada em ontologias, de modo a torná-las disponíveis também para futuros potenciais usuários e autores da Web Semântica. Pode-se notar também que muitas das características identificadas são complexas e susceptíveis a erros, e podem distrair o consideravelmente o usuário/autor durante o processo de produção do documento.

A partir das análises realizadas, identificamos um conjunto de características comuns e habilidades básicas, que, de acordo com o estado da arte, serão exigidas dos autores e usuários desses ambientes de produção de conteúdo para a Web Semântica. São elas:

- familiaridade com técnicas de edição de ontologias;
- extração da informação e anotação semântica manual ou semi-automática;
- categorização da informação e organização do conhecimento;
- redação simultânea de textos e anotação semântica de termos;

- criação de metadados, não necessariamente a partir de padrões;
- checagem de consistência de sintaxe e formato html e outros;
- submissão de novas relações nas ontologias para 2ª opinião.

Os autores de hoje ainda não estão familiarizados com esse tipo de conhecimento, o que nos leva a pensar que a pesquisa em autoria para a Web Semântica está um pouco distante da realidade prática corrente, ou vice-versa, ou seja, as habilidades necessárias para a realização dessas tarefas está restrita a um pequeno grupo de especialistas, ou seja, os próprios engenheiros de ontologias e de conhecimento. No entanto, se os provedores de soluções especialistas para a Web Semântica buscam torná-la acessível para todos em qualquer lugar a qualquer hora, deve-se levar em consideração que esse ambiente deve ser menos complexo do que o constatado nesta pesquisa. O aprendizado obtido com a análise dos sistemas permitiu identificar que os ambientes de autoria analisados estão ainda longe de satisfazer o usuário ou autor de hoje.

### **3.8 Conclusão das análises**

O resultado das análises feitas com o critério definido nesse trabalho revelou o grande potencial das ferramentas de autoria baseadas em ontologias para a representação do conhecimento na Web Semântica. Por ser um assunto novo, o maior desafio em abordar a questão da autoria para a Web Semântica está na ausência de critérios de análise dessas tecnologias, assim como de referenciais teóricos e metodológicos para tanto.

Nos ambientes analisados de autoria baseados em ontologia, aos autores são solicitadas algumas habilidades e um conjunto de técnicas específicas do domínio da ontologia que ainda não fazem parte de suas realidades, tais como: edição de ontologias, anotação semântica, inferência e criação de relações conceituais por meio das ontologias, dentre outras. No entanto, quão aptos estão os autores a usarem essas ferramentas? As novas formas de produção de documentos suportados pela Web Semântica já são uma realidade, mas há também um conjunto de técnicas e habilidades a serem adquiridas de forma a melhorar a representação do conhecimento durante a autoria, de modo a tornar esses sistemas amplamente úteis. Novas tecnologias serão desenvolvidas e novas pesquisas científicas deverão ser desenvolvidas para alcançar esses objetivos.

A partir da abordagem apresentada, levantamos questões importantes e apresentamos algumas respostas ao assunto da autoria baseada em ontologia. O próximo passo desta pesquisa está em desenvolver um protótipo de um ambiente de autoria, conforme proposto no Projeto de Pesquisa de Mestrado da Universidade Vrije de Amsterdam e junto ao Grupo de Pesquisa Arquitetura da Informação da Universidade de Brasília, Brasil (OLIVEIRA, 2004). A base para esse novo ambiente de autoria deve ser formada a partir da definição de requisitos e na integração e novas ou já disponíveis soluções e características, em conjunto com os provedores de soluções, de modo a alcançar o objetivo maior de tornar a Web Semântica disponível para todos e em qualquer lugar, conforme preconizado por Berners-Lee.

**Capítulo 4**  
**Proposta de Arquitetura de Ambientes de Autoria da**  
**Web Semântica**

## **4. Uma proposta de arquitetura de ambientes de autoria para a Web Semântica**

Neste capítulo apresentamos o resultado final da tese, com a proposta de arquitetura de sistemas de informação para ambientes de autoria baseados em ontologias para a Web Semântica. A especificação da arquitetura será apresentada de duas formas: i) especificação do modelo de arquitetura da informação, de acordo com o modelo proposto por LIMA-MARQUES e MACEDO (2005); ii) elaboração do anteprojeto do ambiente, o qual é composto pelos seguintes elementos:

1. especificação de requisitos, segundo a norma IEEE 830 (1998);
2. desenho de interfaces de software;
3. desenho dos diagramas de caso de uso do sistema;
4. desenho do modelo de domínio do sistema;
5. desenho do diagrama de atividades do sistema;
6. desenho do diagrama de seqüência de atividades do sistema;

O modelo de arquitetura para ambientes baseado em ontologias, proposto neste capítulo, leva em consideração os desafios apresentados nas recomendações do W3C com relação a tecnologia independente de dispositivo (W3C, 2003), ou seja, é um modelo genérico. Ele orienta que novos ambientes de autoria para a Web Semântica em geral devem dar suporte a autores na criação de conteúdos e aplicações, assim como considerar novas tecnologias e características, diversas plataformas, protocolos e redes de comunicação, etc. Os autores da Web Semântica devem poder fornecer conteúdo com a mesma facilidade com que é feito hoje na Web, ou seja, um ambiente de autoria deve dar suporte de acesso a uma variedade de dispositivos e mecanismos, para diversos contextos, sem que haja restrição ao tipo de aplicações ou documentos que os autores buscam criar.

Um requisito principal para esses ambientes é com relação a ontologias de contexto ou de domínio: elas devem ser criadas e estar maduras o suficiente para prover aos autores bases de conhecimento consistentes, deixando para eles o menor esforço possível na edição e população da ontologia. Com base nessas premissas, propomos neste capítulo uma arquitetura geral para ambientes de autoria baseada em ontologias para a Web Semântica, um ambiente independente de dispositivo.

O desenho do modelo de arquitetura ora apresentado baseia-se na proposta de LIMA-MARQUES (2005), que consta de um Modelo Genérico de Arquitetura da Informação (Figura 16). O modelo ilustra os processos básicos do ciclo da informação, sobrepostos em três níveis, sendo eles:

- o nível de meta-modelagem;
- o nível de modelagem; e
- o nível de aplicação.

Os três níveis podem ser analisados sob o ponto de vista do modelo clássico da administração, com as categorias operacional-tático-estratégico. Assim, o modelo apresentado pode ser aplicado a ambientes informacionais de qualquer natureza, independentemente do tipo de suporte utilizado, formato, conteúdo ou tipo de informações que o constituem, desde o de uma tradicional biblioteca ao de uma complexa organização, passando pelos ambientes interativos da *Web*, incluindo a *Web Semântica*. O modelo não está, portanto, atrelado a pessoas, organizações ou tecnologias de um ambiente específico (MACEDO, 2005). O tipo de suporte, ou seja, a tecnologia que está por trás da arquitetura, certamente exerce fortes influências no uso do ambiente. Recursos digitais realmente oferecem inúmeras opções de organização e disponibilização de conteúdos, no entanto, os demais suportes não devem ser ignorados, pois continuarão existindo e necessitando de organização para serem recuperados e assim garantir a interoperabilidade do ambiente.

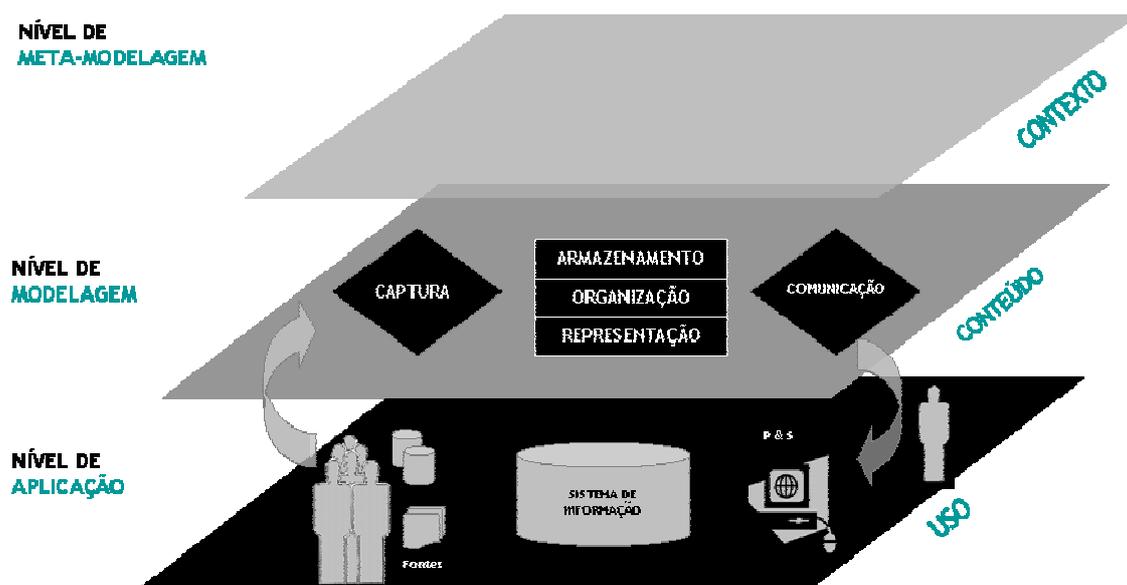


Figura 16: Modelo Genérico de Arquitetura da Informação (Lima-Marques e Macedo, 2005)

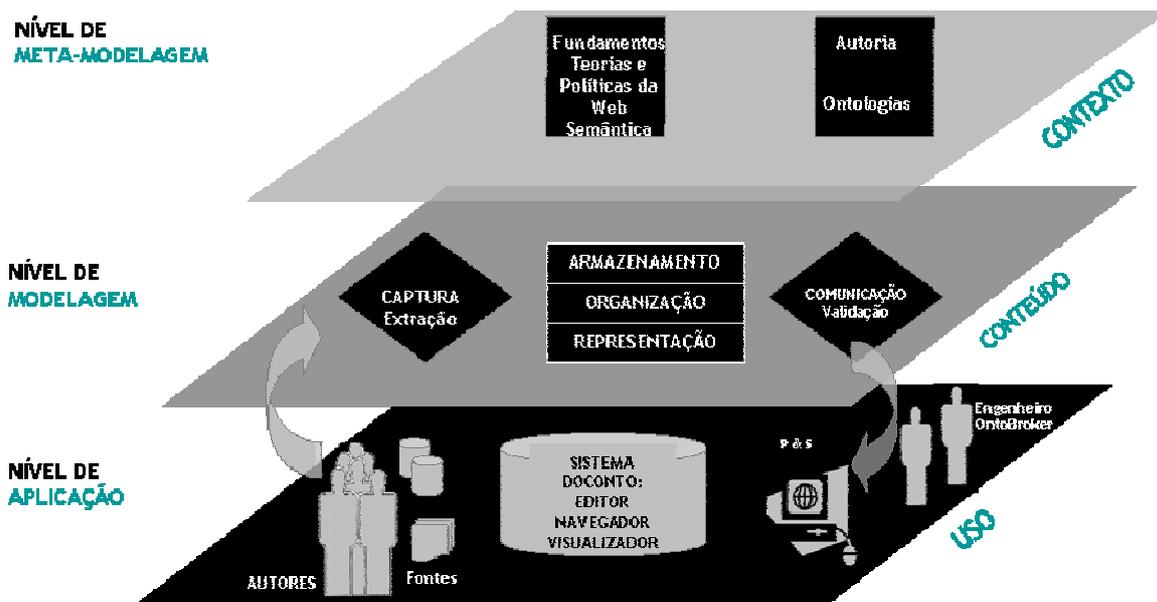
Nesse modelo, o nível meta-modelagem, considerado epistemológico ou estratégico, é o mais alto e menos tangível na representação. É nesse nível que são desenvolvidas as teorias e modelos da área que se busca arquitetar por meio de um ambiente informacional. Aqui é analisado o contexto do ambiente a ser desenhado para representar o arcabouço teórico do trabalho proposto, os quais foram apresentados no Capítulo 2 desta tese.

O nível de modelagem, também tido como um nível científico ou tácito, é onde se definem os modelos relacionados ao contexto. Nele estão contidos os modelos de captura e de tratamento. No modelo de captura, definem-se as diferentes origens dos conteúdos (fontes internas e externas, gestores e outros sistemas) e as características dos conteúdos (natureza, tipologia, formato, suporte), com base no nível estratégico apresentado. No modelo de tratamento, estão contidos três módulos: de representação, de organização e de armazenamento das informações. Essas características também estão descritas na especificação de requisitos, detalhadas no anexo II desta tese.

No Modelo de Comunicação, são considerados os aspectos do ambiente informacional, referentes à recuperação e disseminação das informações, com relação a produtos e serviços que o sistema poderá gerar, com base no nível estratégico definido. A especificação do ambiente de autoria ora proposto baseou-se nesse modelo para tratar da forma com que se dá a interação entre os atores do sistema, com base em requisitos de funcionalidade propostos.

No nível da aplicação estão presentes os elementos tangíveis, ou seja, os sistemas de informação, os usuários, as fontes de informação etc., e onde é viabilizado o uso do ambiente informacional. Aqui se considera o i) uso de ferramentas de tecnologia da informação disponíveis no mercado; ii) definição da infra-estrutura de sistemas e equipamentos, incluindo segurança.

A partir dessa proposta, apresentamos a seguir o modelo de arquitetura para o ambiente de autoria de documentos para a Web Semântica, com seus respectivos níveis:



**Figura 17: Modelo de Arquitetura de Ambiente de Autoria para a Web Semântica**

Buscou-se utilizar esse modelo como macroestrutura para o desenho da arquitetura proposta nesta tese, por ser um modelo que se aplica a diferentes ambientes informacionais de qualquer natureza, independentemente do suporte, ou seja, qualquer contexto constituído por conteúdos e que sirva a uma comunidade de usuários.

A seguir apresentaremos a especificação dos elementos constantes de cada um dos níveis do modelo, considerando-se o nível de meta-modelagem, o nível de modelagem e o nível de aplicação do ambiente.

#### **4.1 Especificação do nível de meta-modelagem da arquitetura**

O nível de meta-modelagem da arquitetura do ambiente de autoria baseado em ontologias partiu do arcabouço teórico apresentado no Capítulo 2 desta tese, e organizado no item 1.3.3 Referencial Teórico e Metodológico. O contexto em questão refere-se à produção de conhecimento na Web Semântica, ou seja, a autoria de documentos com base em ontologias para gerar documentos semanticamente representados.

O contexto representado nessa meta-modelagem tem como base a definição de ambientes de autoria baseados em ontologias, conforme o enunciado proposto nesta tese (OLIVEIRA, 2004):

*“um conjunto de ferramentas de escrita, edição e representação de documentos que interativamente apóiam os usuários (autores), permitindo um melhor acesso, uso e representação semântica do conhecimento durante a autoria, por meio das seguintes tarefas:*

- *fazer anotação semântica de documentos;*
- *criar metadados;*
- *ligar os termos do documento com as ontologias externas;*
- *ligar documentos similares entre si;*
- *transformar citações em links rotulados;*
- *tornar explícita a estrutura retórica de um documento; e*
- *ligar documentos a comentários de revisão de pares.”*

O sistema proposto está baseado na perspectiva cognitiva do processo de escrita, que por sua vez está arraigado na capacidade de interconexão e alternância entre os sub-processos da redação de textos. O modelo cognitivo explicita três modos de visão de documentos em produção: visão de redes de nós que permite o autor colocar as idéias e ligá-las em uma rede de associações (ambiente do editor de texto associado ao ambiente de autoria); visão de estrutura para criar e manipular a estrutura do texto (editores de texto correntes) e visão linear que permite a digitação do texto com o mínimo de interferência. O modelo proposto nesta tese visa associar ambientes de edição de texto com ambientes de ontologia e de descoberta de conhecimento, viabilizando uma plataforma tecnológica que reúne os ambientes de autoria, representação da informação e pesquisa de fontes de informação, conforme descrito adiante, no nível de modelagem e no nível de aplicação dessa arquitetura.

O ambiente tem como base de edição de textos o modelo de hipertextos e hiperímia, onde os autores podem registrar termos e criar *links* entre eles e outros objetos, conforme é feito correntemente em ambientes Web, e que compõem a infra-estrutura para a produção de documentos na Web Semântica.

A linguagem escolhida para a modelagem do ambiente teve como referência os modelos da UML – Unified Modelling Language, e resultou na elaboração dos diagramas de caso de uso e de seqüência de atividades, apresentados no anexo III desta tese, constantes do nível de modelagem do sistema, descrito a seguir.

## **4.2 Especificação do Nível de Modelagem**

O nível de modelagem do ambiente proposto foi elaborado a partir da especificação de requisitos do sistema, apresentada no anexo II desta tese e redigido a partir do padrão da norma IEEE 830 (1998) para especificação de requisitos de sistemas.

A modelagem do sistema foi realizada a partir de dois diagramas principais, representados em UML: diagrama de caso de uso do sistema e diagrama de seqüência de atividades.

As respectivas funcionalidades do ambiente proposto estão distribuídas em categorias, representadas abaixo, e detalhadas na especificação do sistema, constantes no anexo II desta tese:

1. Configuração do Ambiente
  - 1.1 Configuração do ambiente de autoria
  - 1.2 Configuração do ambiente de ontologia
  - 1.3 Configuração do ambiente de descoberta de conhecimento
2. Uso do Ambiente
  - 2.1 Seleção das ontologias para uso
  - 2.2 Uso do ambiente de autoria
  - 2.2 Uso do ambiente de extração da informação
  - 2.3 Uso do ambiente de ontologias
  - 2.4 Uso do ambiente de descoberta de conhecimento.

Os atores do ambiente de autoria proposto são:

**Autor** – pessoa ou grupo de pessoas que produz um documento.

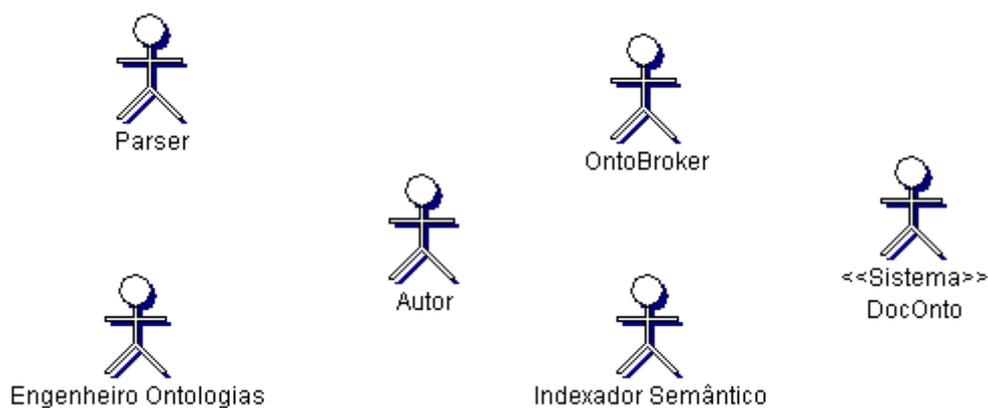
**Engenheiro de ontologias** – pessoa responsável por verificar a consistência da ontologia do documento criada pelo autor, a associação dessa com as ontologias existentes e a criação de novas relações ontológicas sugeridas pelo autor.

**Ontobroker** – pessoa ou empresa que realiza trabalhos de elaboração ou hospedagem de ontologias de domínios.

**Parser** – mecanismo eletrônico que realiza a categorização de entidades léxicas a partir de uma análise sintática automática.

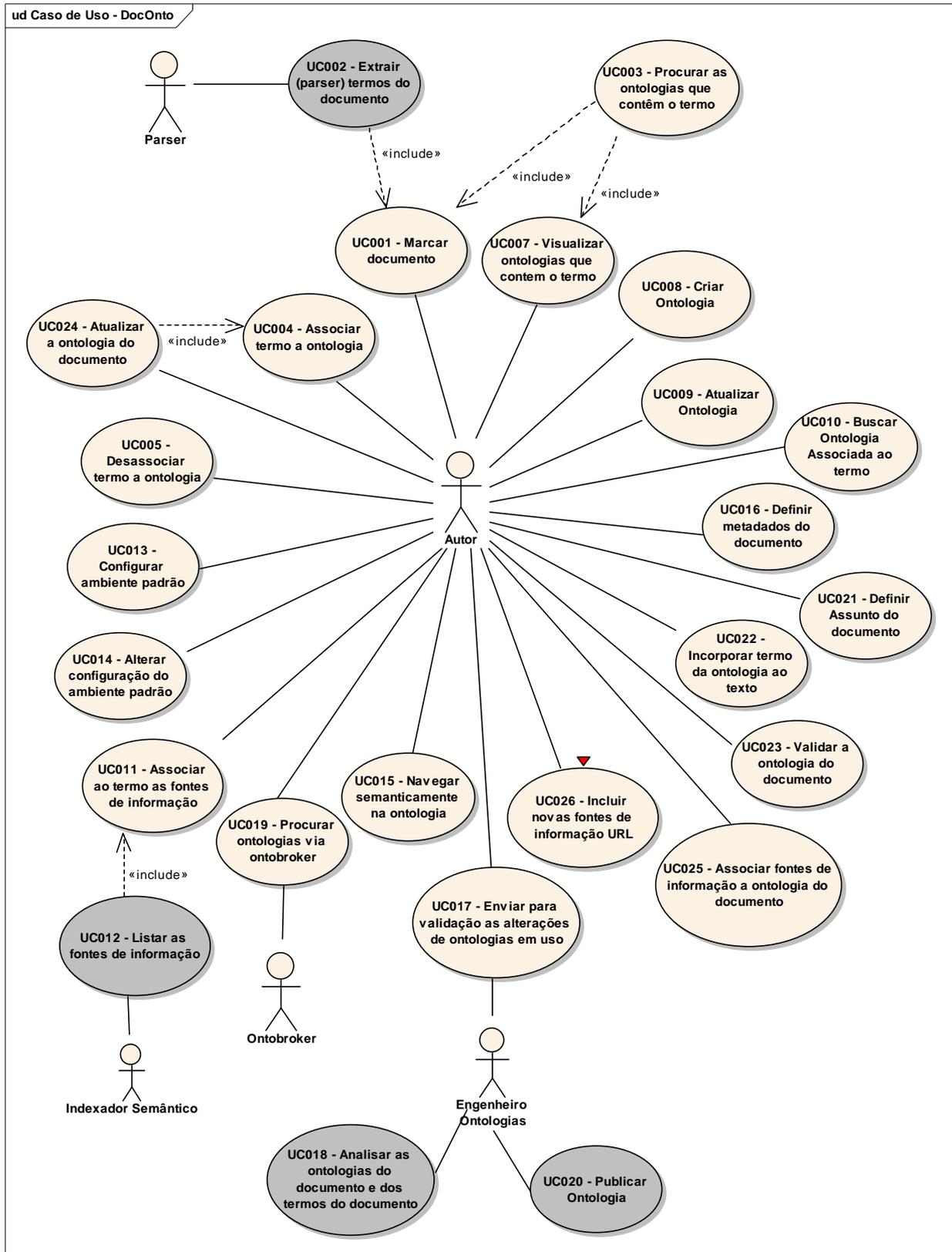
**Indexador semântico** – mecanismo eletrônico que realiza a associação de fontes de informação do ambiente de descoberta com as instâncias reconhecidas na ontologia.

**Sistema DocOnto** – nome da solução proposta de ambiente de autoria baseado em ontologias para a Web Semântica.

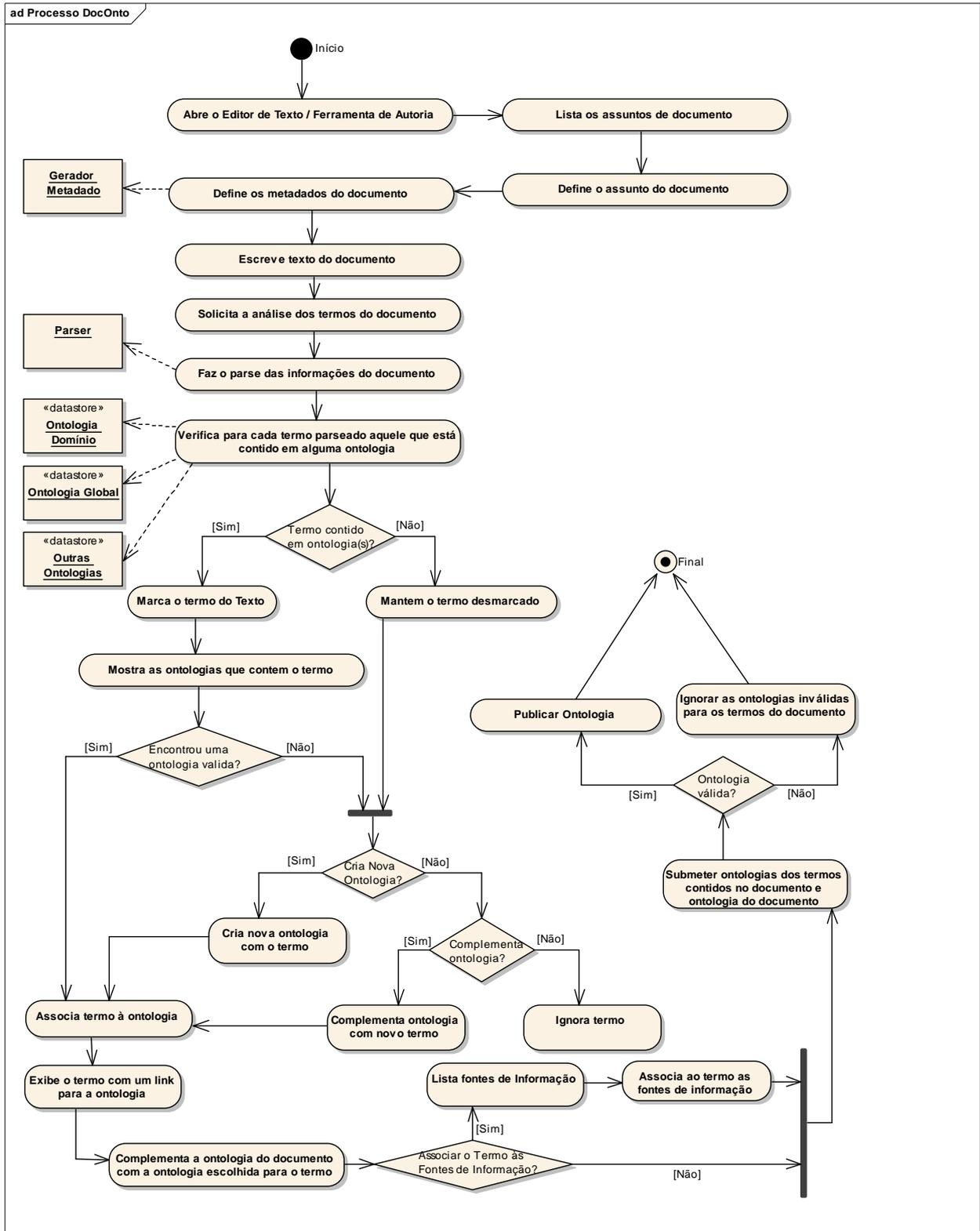


Apresentamos a seguir o diagrama de caso de uso do sistema e o diagrama de seqüência de atividades, com as funcionalidades respectivas desempenhadas pelos atores do sistema. As funcionalidades estão descritas em detalhe no anexo II desta tese.

## Diagrama de caso de uso do ambiente de autoria



# Diagrama de atividades do sistema



Os módulos de representação, organização e armazenamento da informação do modelo de arquitetura do ambiente de autoria foram elaborados com base nas recomendações do W3C para a autoria de documentos na Web Semântica com o uso de ontologias. Desta forma, detalhamos, no anexo II desta tese, as funcionalidades que o sistema apresenta e quais requisitos devem ser atendidos, de modo a alcançar a conformidade desse ambiente com as recomendações que o W3C sugere para esse tipo de tecnologia, assim divididas:

- requisitos para autoria de documentos em XML.
- requisitos para uso de ontologias em OWL;
- requisitos para autoria de documentos compostos;
- requisitos para acessibilidade do ambiente de autoria;
- requisitos para autoria independente de dispositivos.

### ***4.3 Especificação do Nível de Aplicação da Arquitetura***

O ambiente de autoria baseado em ontologia para a Web Semântica destina-se ao uso de autores de documentos eletrônicos em geral, ou seja, qualquer pessoa que se utiliza de ambientes informatizados para a elaboração de um texto e suas respectivas atividades de edição e publicação, levando em consideração as potencialidades que a Web Semântica oferece.

O sistema ora especificado é destinado ao uso em qualquer ambiente de produção de conhecimento, especialmente para a autoria de textos com representação do conhecimento a partir de ontologias. O objetivo principal do sistema é prover aos autores uma infra-estrutura ontológica para que possam localizar conceitos, interligar os conceitos com fontes de informação, representar os metadados do documento de forma automática ou semi-automática e permitir a navegação semântica e incrementar a recuperação da informação em meio eletrônico, durante a autoria do documento.

Conforme apresentado no Capítulo 2, um ambiente de autoria baseado em ontologias depende da integração de diversas soluções. Nesta seção é apresentada a relação entre a tecnologia proposta e outras tecnologias complementares. O

ambiente de autoria de documentos baseado em ontologias é uma solução que se integra com outras tecnologias de forma interoperável, visando atingir os objetivos propostos. Conforme a Figura 18, o modelo proposto visa integrar as seguintes funcionalidades:

- a) um ambiente de autoria e de edição de textos;
- b) um ambiente de extração da informação;
- c) um ambiente de anotação e marcação semântica;
- d) um ambiente de ontologias;
- e) um ambiente de descoberta de conhecimento.

Apresentamos a seguir o modelo genérico do ambiente de autoria baseado em ontologias que, em linhas gerais, representa as seguintes funcionalidades: ao anotar automaticamente os termos de um documento e relacioná-los às ontologias, os documentos serão gerados com uma ontologia específica e com *hiperlinks* para geração automática de metadados, indexação e navegação conceitual de recursos associados. O ambiente permitirá ao autor produzir novos documentos, indexar, fazer *hiperlinks*, publicar e recuperar documentos existentes, com base em relações semânticas precisas oferecidas pela ontologia do documento que foi criada automaticamente ou semi-automaticamente durante a autoria. Outros requisitos gerais e tão importantes desse ambiente é que ele seja fácil de ser instalado e utilizado, customizado de acordo com as necessidades específicas de cada contexto de produção do conhecimento, considerando a presença de ontologias globais que ajudam na escolha de conceitos de alto nível e que possam também auxiliar na integração e união de ontologias específicas e relacionadas. Um visualizador de ontologias provê uma interface para a identificação de instâncias e a navegação pelos conceitos. O ambiente deve também prover anotação semântica automática de termos e conceitos reconhecidos pelas ontologias, e no caso da anotação não ser feita automaticamente, ou seja, manualmente, uma interface tutorial deve orientar os usuários no desempenho dessa atividade. Deve-se deixar o mínimo de edição de ontologia para os autores, e apenas no caso de novos termos que não são reconhecidos pelas ontologias. No entanto, essa tarefa pode também passar pela avaliação de um engenheiro de ontologia ou de um especialista da área, para que o novo conceito seja de fato inserido na ontologia.

A geração de metadados também deve ser feita automática ou semi-automaticamente, com base nas necessidades específicas de cada usuário, de

acordo com o contexto e padrão de metadados a ser utilizado, como o Dublin Core, por exemplo.

Interoperabilidade com outras tecnologias relacionadas deve ser contemplada para cada tipo de necessidade e contexto temático, assim como com outros processadores de texto externos, editores de ontologias e visualizadores, tesouros, outros geradores de metadados e ferramentas de anotação. É fundamental que estejam de acordo com as recomendações da Web Semântica e suas diretrizes, no entanto, o ambiente deve também coexistir com a antiga Web, ou seja, a Internet de hoje, considerando a necessidade de se interagir com fontes de informação cujo conhecimento não esteja ainda semanticamente representado de maneira formal.

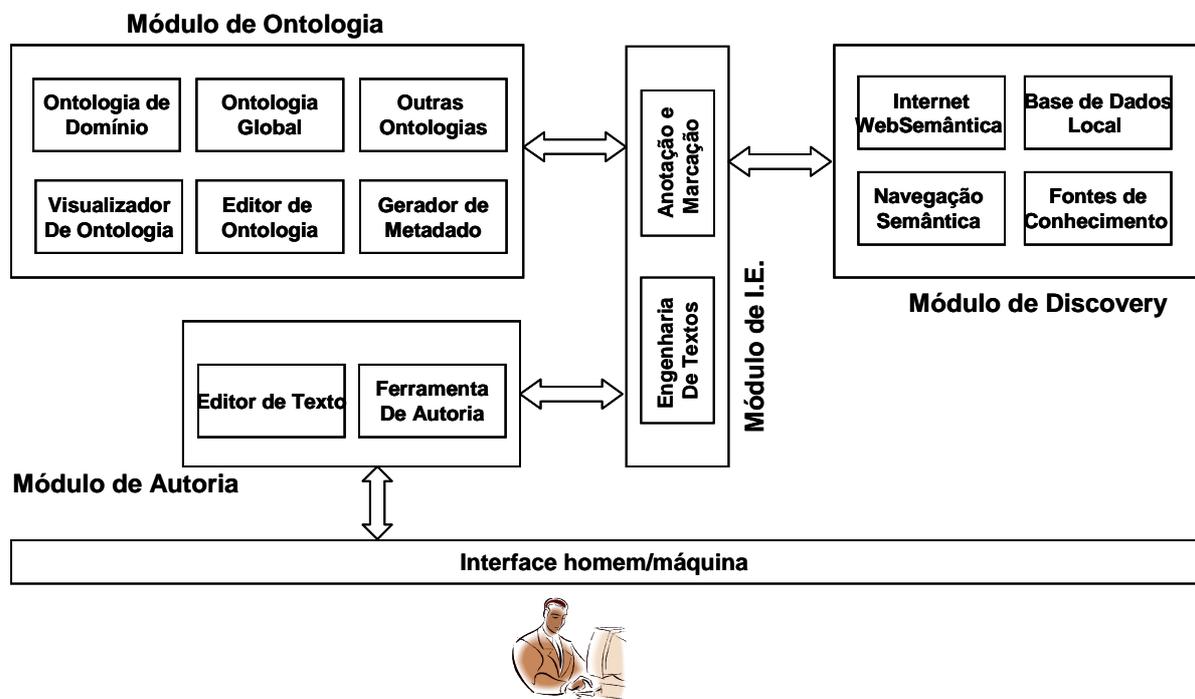


Figura 18 – Modelo geral de arquitetura de ambiente de autoria baseado em ontologia

### **a. Autoria e edição**

- i. Redação de documentos com *links* para as ontologias e recursos de informação.
- ii. Ambiente integrado com editores de textos e ferramentas de autoria existentes.

### **b. Extração da informação**

- i. Extração automática de informação a partir de documentos e bases de dados.
- ii. Anotação e marcação semântica do documento.
- iii. Geração e integração de metadados.

### **c. Visualização de ontologias**

- i. Seleção e validação de conceitos.
- ii. Localização do conceito na árvore e conhecimento (ontologia).
- iii. *Link* visual para ambiente de descoberta.

### **d. Edição de ontologias e metadados**

- i. População da ontologia.
- ii. União entre ontologias de domínio.
- iii. Criação automática de ontologia do documento.
- iv. Criação de metadados do documento a partir de padrões definidos.

### **e. Ambiente de descoberta de conhecimento**

- i. Estrutura para navegação conceitual.
- ii. Criação de trilhas para referência às fontes de informação com *hiperlinks*.

O ambiente é composto por quatro módulos que interagem entre si. A entrada no sistema é feita pela interface homem/máquina que consta de uma aplicação geral que agrega as interfaces de cada um dos módulos do ambiente para uso e navegação. A interface contém 3 ambientes (Figura 19).

- Tela do editor de texto / ferramenta de autoria – com informação extraída e anotada.
- Tela do ambiente de ontologia – com funcionalidade principal de edição e visualização de ontologias.
- Tela do Internet Explorer para navegação em fontes de informação externas.

Consideramos abaixo alguns requisitos gerais que o sistema deve apresentar.

1. Fácil de usar e de configurar, de acordo com as necessidades e habilidades específicas dos usuários.
2. Considerar os diversos tipos de usuários: autores, colaboradores, revisores, engenheiros de ontologias, editores, indexadores e leitores.
3. Interoperabilidade com tecnologias correlatas existentes: processadores de textos, editores e visualizadores de ontologias, tesouros, ferramentas de metadados, anotação e publicação, navegação semântica e ambiente de descoberta de conhecimento.
4. Conformidade com as recomendações e orientações da Web Semântica.
5. Conformidade com as recomendações delineadas no estado da arte apresentado nesta tese.
6. Conformidade com os requisitos identificados na pesquisa preliminar de característica de softwares, realizada nesta tese e apresentada no Capítulo III.
7. Suporte aos ciclos da informação: escrita, edição, revisão, publicação, indexação, recuperação e uso.

Ele é um sistema utilizado para o controle e sugestão de termos digitados em um texto, relacionando-os a uma estrutura ontológica que ajuda os autores a:

- fazer *links* entre os termos do documento e as ontologias de domínio;
- acessar um ambiente de descoberta interligado via ontologias para pesquisa e seleção de outros termos semanticamente relacionados;
- exibir as variações terminológicas que os termos podem apresentar, de modo a suscitar consciência dos usuários de possíveis relações de sinonímia e outras equivalências, de modo a ajudá-los na padronização da representação do texto de acordo com padrões terminológicos de domínios de conhecimento específicos, representados pelas ontologias;
- fazer *links* semânticos entre os documentos e referências com recursos em linha, por meio das ontologias;
- criar automaticamente os metadados do documento a partir de modelos previamente definidos;

- escrever textos em qualquer ambiente institucional, técnico ou científico, onde a ocorrência de ambigüidade deve ser reduzida e uma relação de univocidade entre os termos e conceitos representados seja necessária;
- ter a autonomia de representar semanticamente o seu próprio documento por meio da criação de uma ontologia do documento.

A interface do usuário é composta por três módulos (Figura 19).

- Tela do editor de texto.
- Tela de ontologias (visualização e edição).
- Tela do ambiente de descoberta (navegação).

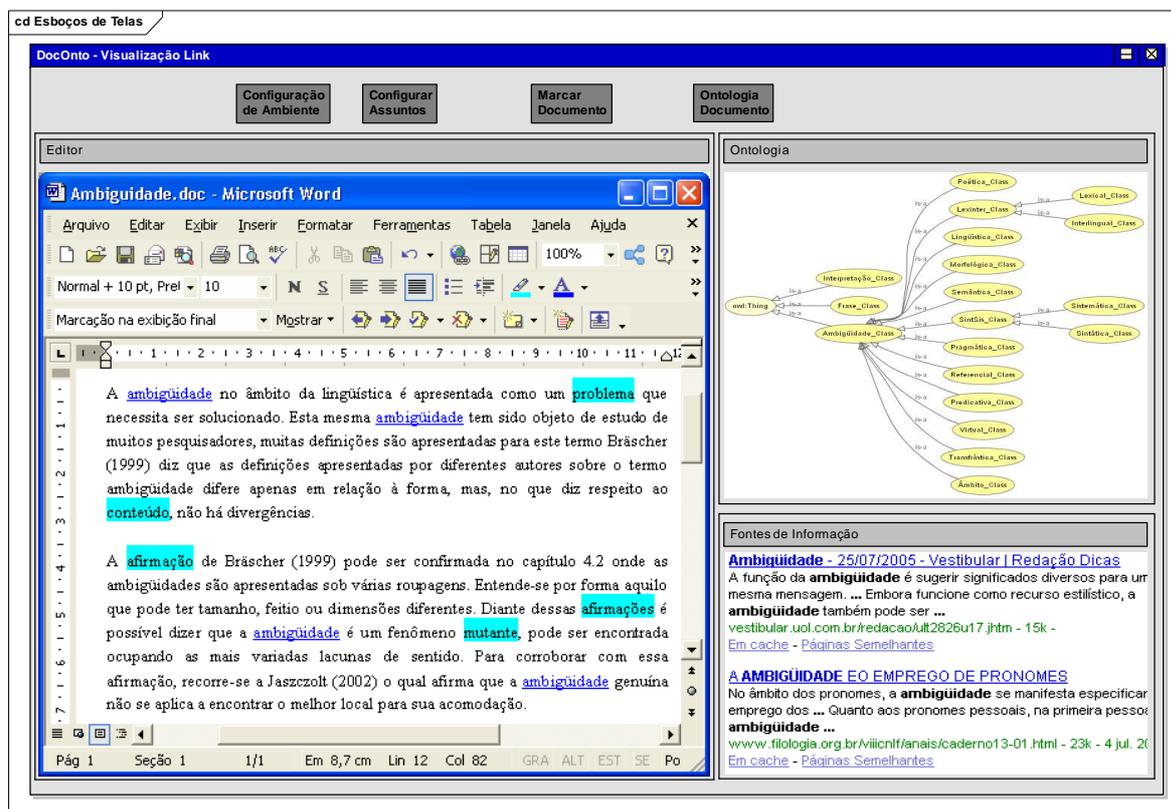


Figura 19 – Tela do Ambiente de Autoria para a Web Semântica

Nesta figura é apresentada a idéia geral do sistema em termos gráficos de interface. O autor terá acesso primeiramente a uma tela de editor de textos comum. Nesse caso, utilizamos o MS Word para ilustrar as funcionalidades básicas de um

editor de textos amplamente utilizado e familiar a usuários de ambientes eletrônicos. O autor tem a liberdade de digitar o texto livremente e fazer associações semânticas antes, durante ou depois da autoria. Se desejar, pode visualizar em tela os três principais ambientes de autoria baseado em ontologias.

Com o desenho dessa interface, apresentado na Figura 19, ilustramos como os três módulos podem ser integrados e disponibilizados ao autor do documento. À esquerda, enquanto digita o texto, os termos reconhecidos na ontologia são anotados (em cores diferentes) e associados à respectiva instância na ontologia (à direita, superior). As fontes de informação referentes aos conceitos digitados aparecem listadas na tela de baixo (à direita).

## **Capítulo 5**

### **Conclusões**

Apresentamos nesta tese o resultado de pesquisa realizado ao longo de empenhados quatro anos de doutoramento, que resultaram em diversas publicações e apresentações científicas sobre o problema de pesquisa ora abordado.

A revisão de literatura sistematizada por essa tese revelou as possibilidades correntes para a autoria de documentos na Web Semântica. Publicamos esse resultado em 2004, durante o congresso internacional de publicações eletrônicas, EIPub (OLIVEIRA, 2004), pelo qual identificamos perante a comunidade internacional um novo estado da arte em autoria para a Web Semântica. Nesse trabalho, mostramos que além de autoria, outras atividades podem ser integradas para se acessar, representar ou gerar documentos ontologicamente representados, com o apoio de tecnologias de extração de informação, engenharia de textos, anotação e marcação semântica de documentos, geração de metadados, visualização ontológica e navegação semântica.

A sistematização conceitual do que é um ambiente de autoria baseado em ontologias para a Web Semântica foi publicado em Hiroshima, Japão, durante o Congresso Internacional da Web Semântica, em 2004 e na Hypertext Conference de 2004, nos Estados Unidos, no fórum Doctoral Consortium (OLIVEIRA, 2004). Definimos pela primeira vez na comunidade científica e da Web Semântica o que é este ambiente, por meio da seguinte proposta conceitual:

*“um conjunto de ferramentas de escrita, edição e representação de documentos que interativamente apóiam os usuários (autores), permitindo um melhor acesso, uso e representação semântica do conhecimento durante a autoria, por meio das seguintes tarefas:*

- *fazer anotação semântica de documentos;*
- *criar metadados;*
- *ligar os termos do documento com as ontologias externas;*
- *ligar documentos similares entre si;*
- *transformar citações em links rotulados;*
- *tornar explícita a estrutura retórica de um documento; e*
- *ligar documentos a comentários de revisão de pares.”*

Concluimos com esse trabalho que de fato novas tecnologias e tendências estão ao dispor da comunidade científica internacional para o desenvolvimento de pesquisas que contribuam para a fundamentação teórica do assunto e busquem soluções práticas para viabilizar uma representação semântica do conhecimento na presente era da informação.

Nessa ocasião, também publicamos a proposta metodológica para se atingir os resultados desta tese (OLIVEIRA, 2004), permeando a coleta e análise de requisitos de ferramentas que apresentam as características acima citadas, considerando-se que, para tanto, um novo critério de análise precisava ser desenvolvido para se atingir esse objetivo. Essa atividade revelou que ainda é necessário desenvolver ferramentas à luz das teorias e recomendações que regem a Web Semântica, assim como considerar a preparação dos potenciais futuros autores desses ambientes a adquirir novas habilidades, como por exemplo, a representação semântica da informação por meio das ontologias.

Concluimos nesta tese que de fato a Web Semântica oferece um novo e promissor paradigma de produção de conhecimento baseado em ontologias, por meio de tecnologias de edição de textos e de representação da informação em meio eletrônico. O processo de autoria de documentos está diante de um cenário inédito de possibilidades que, se integradas, permitem a realização de atividades jamais vistas anteriormente na história do ciclo da informação, em especial a etapa da produção de informação. Assim, buscou-se explicitar quais são as funcionalidades desse novo cenário, o que nos permitiu organizá-las e também levantar novas funcionalidades com base na pesquisa de ferramentas existentes e na proposta de novos requisitos específicos para um ambiente de autoria. Para alcançar esse objetivo, propusemos um critério de análise de ambientes de autoria baseados em ontologia, que poderá ser utilizado por diversos setores e grupos, quando do desenvolvimento de novas soluções para a Web Semântica. Dada a carência desse tipo de abordagem na literatura corrente, esperamos que a organização, análise e proposta do critério de análise ora apresentado sejam úteis na orientação desses trabalhos. O resultado final da análise de requisitos das ferramentas existentes para autoria em diversas plataformas e funcionalidades demonstrou ser recente essa preocupação. É necessário que instituições de pesquisa e desenvolvimento de sistemas possam estar apoiadas a desenvolver ambientes de autoria que façam uso

das potencialidades da Web Semântica e que estejam de acordo com as recomendações do W3C e da comunidade de engenharia de software.

O modelo de ambiente arquitetado nesta tese demonstrou ser possível integrar soluções correntes que facilitam a autoria, a representação ontológica da informação e a navegação semântica. Publicamos esse resultado no congresso EIPub 2006 (OLIVEIRA, 2006), no qual aplicamos os critérios de análise desenvolvido à luz das recomendações da Web Semântica, assim como critérios correlatos, revelando que as ferramentas hoje existentes ainda não estão prontas para prover soluções de acordo com as bases fundamentais da Web Semântica. Portanto, neste trabalho apresentamos um modelo genérico de arquitetura que atenda ao atual paradigma de representação do conhecimento para a Web Semântica, onde também revelamos que é possível ter em um mesmo ambiente os processos de autoria, representação e pesquisa de informação. Para tanto, um esforço de sistematização de requisitos torna-se eminente para suscitar consciência de desenvolvedores de tecnologias, quanto às adequações que se torna necessário considerar para que se atinja o objetivo primordial de tornar a Web Semântica acessível a todos e em qualquer lugar, conforme preconizado por Tim Berners-Lee.

Assim, atingindo esse objetivo, apresentamos no anexo II desta tese uma completa especificação de requisitos com base em padrões internacionais, pela qual foi possível realizar a modelagem do então proposto ambiente de autoria baseado em ontologias para a Web Semântica. A solução proposta integra três ambientes: um editor de texto, um ambiente de ontologias e um ambiente de descoberta de conhecimento. Por meio da identificação semântica dos termos digitados em um documento, o sistema pode associar os termos a classes e instâncias da ontologia e assim criar uma lista de fontes de informação correlatas para pesquisa e navegação semântica. Propomos também, como consequência dessa ação, que o documento representado possui em si uma identidade ontológica que pode ser utilizada no futuro para a indexação de documentos ontologicamente representados em sistemas de informação. Sugerimos aqui a continuidade de trabalhos futuros especificamente sobre esse problema da “representação ontológica da informação”, a qual poderá beneficiar em grande profundidade as atuais visões de tratamento, representação e recuperação da informação no âmbito da Ciência da Informação.

As publicações acima referidas, assim como a visão de ambiente de autoria para a Web Semântica apresentados nesta tese estão sendo utilizados

correntemente pelo grupo de pesquisa Arquitetura da Informação, do CID/UnB, coordenado pelo Prof. Dr. Mamede-Lima Marques, orientador desta tese. O trabalho foi utilizado como base para a elaboração de dissertação de mestrado de SILVA (2006) que propõe um modelo ontológico para a resolução de ambigüidades na autoria de documentos. O modelo de arquitetura do ambiente foi utilizado como abordagem conceitual de referência para a dissertação do aluno de mestrado Daniel Parente, que busca criar um arcabouço teórico para a representação ontológica da informação para a diminuição de ambigüidades na recuperação e interpretação da informação.

O futuro deste trabalho está em desenvolver e prototipar o ambiente ora especificado, por meio de uma metodologia específica, a qual está proposta como tema de dissertação de mestrado do departamento de Inteligência Artificial da Vrije Universiteit de Amsterdam, sob a orientação do Prof. Dr. Frank van Harmelen (OLIVEIRA, 2004).

Entendemos assim que está diante de nós um caminho próximo de aplicabilidade e desenvolvimento do ambiente de autoria ora proposto. Com esta tese, contribuimos para o desenvolvimento de soluções que estejam de acordo com os princípios e potencialidades da Web Semântica. É fundamental que se tragam esses assuntos para serem discutidos no âmbito da Ciência da Informação, de modo a solidificar as teorias que regem esses assuntos, garantir a satisfação dos usuários desses ambientes, ao mesmo tempo em que se busca o sucesso da representação da informação em meios eletrônicos, o que representa hoje um grande problema e desafio para essa área, em conjunto com aportes da Ciência da Computação e da lógica matemática.

## 6. Bibliografia

- ALBERINK, M et al. Clustering semantics for discourse generation. The Hypermedia and the Semantic Web Workshop at the **Fourteenth Conference on Hypertext and Hypermedia (Hypertext 2003)**, 2003.
- ALMEIDA, Maurício, BAX, Marcello. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. In **Ciência da Informação**. Vol. 32, N°3. 2003.
- ANDREASEN, T. et al. Content-based text querying with ontological descriptors. **Data & Knowledge Engineering**. Elsevier Science Publishers B. V, Amsterdam, 2004.
- ANGELE, J. et al. **EFFORT - Evaluation framework for ontologies and related technologies**. 2003. Disponível em <<http://www.ontoprise.de/documents/effort-ekaw.pdf>>.
- \_\_\_\_\_, SURE, Y.: **Whitepaper: Evaluation of ontology-based tools**. Excerpt from the IST-2001-29243 Report, OntoWeb. D1.3. Tools (2001). Disponível em: <[http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/eon2002\\_whitepaper.pdf](http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/eon2002_whitepaper.pdf)>.
- ANTONIOU, Grigoris; HARMELEN, Frank van. **A Semantic Web Primer**. Fundamentals of the technologies inherent to the Semantic Web applications. MIT Press.2004.
- AROYO, L. et al. **A layered approach towards domain authoring support**. In Proceedings of ICAI 2002 (Las Vegas, US) CSREA Press. Disponível em: <<http://citeseer.ist.psu.edu/aroyo02layered.html>>.
- BÄCK, A. et al. **Semantic Web technologies in knowledge management**. In proceedings EI Pub 2003, Portugal.
- BARRETT, Edward (ed.) **Text, context, and hypertext: writing with and for the computer**. Cambridge : The MIT Press, 1988.
- BASIL, V. et al. **Experimentation in software engineering**. IEEE Transactions in Software Engineering, SE-12(7):733-743. No. 7, July 1986.
- BAX, Marcello Peixoto. Introdução às linguagens de marcas. **Ciência da Informação**, vol.30, no. 1, 2001. Disponível em: <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=256>>.
- BEREITER C., SCARDAMALIA. M. **The psychology of written composition**. Hillsdale, NJ: Erlbaum. 1987.
- BERNERS-LEE, Tim. **Weaving the Web**: the original design and ultimate destiny of the World Wide Web by its inventor. Harper San Francisco. 1<sup>st</sup> Edition, 1999.
- \_\_\_\_\_, J. HENDLER, O. Lassila. The Semantic Web. **Scientific American**, May 2001.
- BOLTER, Jay David. **Writing space**: the computer, hypertext, and the history of writing. Lawrence Erlbaum Associates, Hillsdale, 1991.
- BONTCHEVA, K. et al. **Evolving GATE to meet new challenges in language engineering**. Natural Language Engineering. 10 (3/4), pp. 349-373. 2004.

- BORGMAN, Christine L. **From Gutenberg to the global information infrastructure: access to information in the networked world.** Cambridge : The MIT Press. 2001.
- BOSTOCK, Stephen. **An evaluation of three authoring tools: PC-CAI. Authorware and Toolbook,** 1994. Disponível em: <[http://www.keele.ac.uk/depts/cs/Stephen\\_Bostock/docs/authass.htm](http://www.keele.ac.uk/depts/cs/Stephen_Bostock/docs/authass.htm)>.
- BRACHMAN, R. et al. Knowledge representation, connectionism and conceptual retrieval. Em **Annual ACM Conference on Research and Development in Information Retrieval** Proceedings of the 11th annual international ACM SIGIR. ACM Press New York, NY, USA . 1988.
- BROCKMANN, R. John. **Writing better computer user documentation.** John Wiley & Sons, Inc. New York, NY, USA . 1990.
- BROWN, J.S., DUGUID, P. The social life of documents. **First Monday**, vol 1, 1996. Disponível em: <<http://www.firstmonday.dk/issues/issue1/documents>>.
- BUSH, V. 1945. *As we may think.* **The Atlantic Monthly**, 1945. Disponível em: <<http://www.theatlantic.com/doc/prem/194507/bush>>.
- CARACCILOLO, C, et al. Towards scientific information disclosure through concept hierarchies. In J. A. Carvalho, A. Huebler, A. A. Baptista (eds.), Proc. of the **6th International ICCI/IFIP Conference on Electronic Publishing (ELPUB02)**, Karlovy Vary (Czech Republic), 2002.
- CARR, L., et al. Towards a knowledge-aware office environment. In **Proceedings of 14th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2004)**, Whittlebury Hall, Northamptonshire, UK. 2004. Disponível em: <<http://eprints.ecs.soton.ac.uk/archive/00009458/>>.
- \_\_\_\_\_. The case for explicit knowledge in documents. ACM: **DocEng'04**, October 28-30, Milwaukee, EUA. 2004.
- \_\_\_\_\_, HALL, Wendy, et al. Conceptual linking: ontology-based open hypermedia. In **WWW10**, Hong Kong, ACM. 2001.
- \_\_\_\_\_, ROURE, David de et al. The distributed link service: a tool for publishers, authors and readers. In **WWW4**, 1995. Disponível em: <<http://www.w3.org/Conferences/WWW4/Papers/178/>>.
- CHEN. **An ontology-based intelligent authoring tool**, 1998. Disponível em: <<http://www.ei.sanken.osaka-u.ac.jp/pub/wendy/wendy-icce98.pdf>>.
- CIMIANO, Philipp, et al. **Towards the self-annotating Web.** 2004. Disponível em: <<http://www.www2004.org/proceedings/docs/1p462.pdf>>.
- CIRAVEGNA, Fabio. et al. Document annotation via adaptive information extraction. Poster at the **25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.** Tampere, Finland. 2002.
- COCH, J. Evaluating and comparing three text-production techniques. Proceedings of the **16th Conference of Computational Linguistics**, Coling 96: 249-254, Copenhagen, Denmark, 1996.
- COOMBS, James, RENEAR, Allen et al. **Markup systems and the future of scholarly text processing.** 1987. Disponível em: <<http://xml.coverpages.org/coombs.html>>.
- CORCHO, Oscar, SURE, York . Evaluation of Ontology-based Tools. Proceedings of the 2nd International Workshop on Evaluation of Ontology-based Tools EON2003. **2nd International Semantic Web Conference ISWC 2003**, 20th October 2003, Sanibel Island, Florida, EUA.

CRISTEA, Alexandra. Evaluating Adaptive Hypermedia Authoring while teaching adaptive systems. **SAC'04** ACM, Nicosia, Cyprus. 2004. Disponível em: <<http://www.wis.win.tue.nl/~acristea/HTML/Minerva/papers/CristeaSAC04CameraReadyLast+2give.pdf>>.

\_\_\_\_\_. et al. **Evaluation of MOT, an AHS authoring tool**: checklist and a special evaluation class. 2003. Disponível em: <<http://www.wis.win.tue.nl/~acristea/HTML/Minerva/papers/CATE-cristea-mooij.doc>>.

DAVIS, Alan M. **Software requirements**: objects, functions, and states. Prentice Hall, 1993.

DELLA VALLE, E. BRIOSCHI, M. Towards a semantic enterprise information portal, In proceedings of **Workshop on Knowledge Management and the Semantic Web at K-CAP'03**, 2003.

DENNY, Michael. **Ontology building**: a survey of editing tools. 2002. Disponível em: <<http://www.xml.com/pub/a/2002/11/06/ontologies.html>>

\_\_\_\_\_. **Ontology Tool Survey**, Revisited . O'Reilly Xml.com, 2004. Disponível em: <<http://www.xml.com/pub/a/2004/07/14/onto.html>>.

DINGLEY, Andy. **Authoring**: in and out of the real world. Invited workshop at *MesMuses 2003*, Florence, 2003.

\_\_\_\_\_; SHABAJEE, Paul. **Today's authoring tools for tomorrow's Semantic Web**. Presented at Museums and the Web 2002, Boston 2002.

DINGLI, Alexiei, CIRAVEGNA, Fabio et al. **Automatic semantic annotation using unsupervised information extraction and integration**. ACM. 2000.

DOMINGUE, John , DZBOR, Martin , MOTTA, Enrico. Semantic layering with Magpie. In S Staab R Studer (ed) **Handbook on Ontologies in Information Systems**, Springer Verlag, 2003.

DORNER, Jane. **Authors and information technology**: new challenges in publishing, in M. Sharples (ed.), *Computers and Writing: issues and implementations*. Dordrecht: Kluwer. 1992.

DUBLIN CORE. **The Dublin Core Metadata**, 2004. Disponível em: <<http://dublincore.org>>.

DYMETMAN, Marc. **Text authoring, knowledge acquisition and description logics**. 2003.

FAABORG, A.,LAGOZE, C. Semantic Browsing. in Lecture Notes in **Computer Science**, vol. 2769. Trondheim, Norway: Springer-Verlag, 2003, pp. 70-81.

FENTON, Norman E., NEIL, Martin. Software metrics: roadmap. In proceedings of the conference on **The future of Software engineering**. ACM Press, 2000. Disponível em: <<http://doi.acm.org/10.1145/336512.336588>>.

FLOWER, L, HAYES, J. R. **Problem-solving strategies and the writing process**. *College English*, 39, 449-461. 1977.

\_\_\_\_\_. **A cognitive process theory of writing**. *College Composition and Communication*, 32 (1), 365-387,1981.

\_\_\_\_\_. The dynamics of composing: making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), **Cognitive processes in writing: An interdisciplinary approach** (pp. 31-50). Hillsdale, NJ: Erlbaum. 1980.

FLUIT, C., SABOU, M., HARMELEN, F. Supporting user tasks through visualization of light-weight ontologies. In **Handbook on Ontologies in Information Systems**, Springer-Verlag, 2003.

FOUCAULT, Michel. **A Arqueologia do Saber**. Petrópolis, Vozes, 1972.

GARCIA-CASTRO, Raúl. **Benchmarking ontology technology**. In Deliverable KnowledgeWeb Project, chapter 4, 2004.

GEEST, Thea van der. Professional writing studied: author's accounts of planning in document production processes. Chapter 1. *in* SHARPLES, Mike, van der GEEST, Thea. **The new writing environment: writers at work in a world of technology**. London: Springer-Verlag, 1996.

GEURTS, J. et al. **Towards ontology-driven discourse: from semantic graphs to multimedia presentations**. 2003. Disponível em: <<http://homepages.cwi.nl/~media/publications/iswc2003.pdf>>.

GIBBONS, M. et al. **The New Production of Knowledge: The dynamics of science and research in contemporary societies**, Sage, London, 1994.

GIL, Yolanda, RATNAKAR, Varun. TRELIS: an interactive tool for capturing information analysis and decision making, In Proceedings of the 13th **International Conference on Knowledge Engineering and Knowledge Management**, Siguenza, Spain, October 1-4, 2002. Disponível em: <<http://www.isi.edu/ikcap/trellis/>>.

\_\_\_\_\_. Trusting Information Sources One Citizen at a Time. **ISWC'02 – International Semantic Web Conference**, Sardinia, 2002.

GÓMEZ-PÉREZ, Asuncion. **Criteria to verify knowledge sharing technology KSL**. Stanford University. 1995. Disponível em: <[http://www-ksl.stanford.edu/KSL\\_Abstracts/KSL-95-10.html](http://www-ksl.stanford.edu/KSL_Abstracts/KSL-95-10.html)>.

\_\_\_\_\_. **Some Ideas and Examples to Evaluate Ontologies**. Technical Report #KSL-94-65, Knowledge Systems Laboratory. Stanford University, 1994.

\_\_\_\_\_, et al. **State of the art in ontologies from the Semantic Web perspective**. Esperanto Deliverable D1.1, 2002.

GREENBERGER, M. The Computers of Tomorrow. **Atlantic Monthly**. Maio de 1964.

GRUBER, T. **Towards principles for the design of ontologies used for knowledge sharing**. Technical Report KSL 93-04. Knowledge Systems Laboratory. Stanford University. CA. 94305.1993.

GRUBER, T. **What is an ontology**, 1996. Disponível em <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.

GUARINO, N. Formal Ontology and Information Systems. In N. Guarino, editor, Proceedings of the **1st International Conference on Formal Ontologies in Information Systems, FOIS'98**, Trento, Italy, pages 3-- 15. IOS Press, Junho 1998.

HANDSCHUH, S. et al. On deep annotation. In Proceedings of the **12<sup>th</sup> International World Wide Web Conference. WWW2003**, Budapest, Hungary. ACM, 2003.

\_\_\_\_\_. S-CREAM – Semi-automatic CREAtion of Metadata. In Proceedings of the 13th **EKAW'02**, 2002.

\_\_\_\_\_. Authoring and annotation of Web pages in CREAM. In Proceedings of the **WWW2002 – Eleventh International WWW Conference**, Hawaii, USA, Maio 2002. Disponível em: <<http://annotation.semanticweb.org/ontomat/index.html>>.

HANSEN, Brad. **The dictionary of multimedia terms and acronyms**. Chicago : FD, 1998.

HARMELEN, Frank van. How the Semantic Web will change KR: challenges and opportunities for a new research agenda. **The Knowledge Engineering Review** vol 17, no 1, 2002.

- HARMSZE, F. et al. A modular structure for electronic scientific articles. Conferentie Informatiewetenschap 1999. CWI, Amsterdam, 1999. In: P. de Bra and L. Hardman (eds). **Computing Science Reports**. Technische Universiteit Eindhoven. Report 99-20. pp. 2-9. Disponível em: <<http://www.cwi.nl/~lynda/WGI/info-wet1999/proceedings>>.
- HASELKORN, Mark P. **The future of “writing” for the computer industry**. In Barrett, Edward (ed.) Text, context, and hypertext: writing with and for the computer. Cambridge : The MIT Press, 1988.
- HAYASHI, Y et al. **A multiple view authoring tool for modeling training materials**. 1999. Disponível em: <<http://www.ei.sanken.osaka-u.ac.jp/pub/hayashi/hayashi-aitr05.pdf>>
- HAZARI, S. I. **Evaluation and selection of Web course management tools**. Disponível em: <<http://www.sunilhazari.com/education/webct>>.
- HEFLIN, J. e HENDLER, J. Dynamic Ontologies on the Web. In **Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000)**. AAAI/MIT Press, Menlo Park, CA, 2000. pp. 443-449.
- HENDLER, J. et al. Integrating applications on the Semantic Web. **Journal of the Institute of Electrical Engineers of Japan**, Vol 122(10), October, 2002, p. 676-680.
- IANNELA, R.; WAUGH, A. **Metadata: enabling the Internet**. Disponível em: <<http://archive.dstc.edu.au/RDU/reports/CAUSE97>>. Acesso em: fev. 2003.
- IEEE **Recommended Practice for Software Requirements Specifications**, 1998. Disponível em: <<http://www.dcc.ufmg.br/~rodolfo/es-1-03/IEEE-Std-830-1998.pdf>>.
- ISO/IEC 9126-1. **Software engineering – Product Quality – Part 1 Quality model 2002**.
- \_\_\_\_\_. 14598-1. **Information Technology – Evaluation of Software Products – Part 1 General Guide**, 1998.
- KALYANPUR, A et al. **SMORE - Semantic Markup, Ontology, and RDF Editor**. 2005. Disponível em: <<http://www.mindswap.org/papers/SMORE.pdf>>.
- KDE Documentation. **Fundamentos do UML**. Capítulo 2. Disponível em: <[http://docs.kde.org/stable/pt\\_BR/kdesdk/umbrello/uml-basics.html](http://docs.kde.org/stable/pt_BR/kdesdk/umbrello/uml-basics.html)>. Acessado em dezembro 2005.
- KHUN, Thomas. **The structure of scientific revolutions**. The University of Chicago Press, 1983.
- KILGOUR, Frederick G. **The evolution of the book**. New York : Oxford University Press.1998.
- KIM, Henry M. **Predicting how ontologies for the Semantic Web will evolve**. *Communications of the ACM*, Vol. 45, No. 2, pp. 48-54. 2002.
- KING, M. Living up to standards. In Proceedings of the **EACL 2003 Workshop on Evaluation Initiatives in Natural Language Processing**, Budapest, Hungary, 2003.
- KIRCZ, J. **Modularity: the next form of scientific information presentation?** 1988. Disponível em: <<http://www.kra.nl/Website/Artikelen/Jdoc98.htm>>.
- KITCHENHAM, B. **Desmet: a method for evaluating software engineering tools**. University of Keele, 1996. Disponível em: <<http://www.keele.ac.uk/depts/cs/se/e&m/tr9609.pdf>>.
- KLEIN, M. **Change management for distributed ontologies**. PhD Thesis. Vrije Universiteit, Amsterdam 2004.

KOIVUNEN, M.R., SWICK, R. Metadata based annotation infrastructure offers flexibility and extensibility for collaborative applications and beyond. In: Proceedings of the **KCAP 2001 Workshop on Knowledge Markup and Semantic Annotation**, Victoria, British Columbia, 2001.

LAMPOR, Leslie. **LaTeX: A Document Preparation System**. Second edition. Addison-Wesley, 1994.

LANDOW, George P. **Hypertext 2.0: the convergence of contemporary critical theory and technology**. London : The Johns Hopkins University Press, 1997.

LAWRENCE, J. D., HARRISON, I.W., RODRIGUEZ, A. C. Capturing analytic thought. In: Proceedings of the First **International Conference on Knowledge Capture K-CAP 2001**, Victoria, British Columbia, 2001.

LE COADIC, Yves François. **A Ciência da Informação**. Brasília: Briquet de Lemos, 1996. 119p.

LERNER, Fred. **The story of libraries: from the invention of writing to the computer age**. New York : Continuum. 2001.

LIMA-MAQUES, Mamede; MACEDO, Flávia. Arquitetura da informação: base para a gestão do conhecimento. In: TARAPANOFF, Kira (Org.) **Gestão da informação e do conhecimento em organizações**. São Paulo, 2005, p. 177-192.

LOC – Library of Congress. **MARC Standards**. 2005. Disponível em: <http://www.loc.gov/marc/>.

MACEDO, Flávia. **Arquitetura da Informação: aspectos epistemológicos, científicos e práticos**. Dissertação de Mestrado, Universidade de Brasília, junho de 2005.

MACKNIGHT, Carol & BALAGOPALAN, Santosh. **An evaluation tool for measuring authoring system performance**. ACM Vol 32, no. 10. 1989.

MACORATTI, José C. **UML: Conceitos Básicos II**. Disponível em: [http://www.macoratti.net/vb\\_uml2.htm](http://www.macoratti.net/vb_uml2.htm). Acessado em dezembro 2005

MARKANTONATOU, S. et al. An authoring tool for controlled modern greek. Proceedings of **2nd Hellenic Conference on AI - SETN-02**, I.P. Vlahavas and C.D. Spyropoulos (eds), 2002, pp. 165-176. Disponível em: <http://iit.demokritos.gr/skel/Ellogon/>.

MICROSOFT. **Visão geral do Word 2003**: o processador de texto do Microsoft Office. 2003. Disponível em: <http://www.microsoft.com/brasil/office/word/overview.asp>.

MONACO, James. **Dictionary of new media**: the new digital world of video, audio, and print. New York : Harbor Electronic Publishing, 1999.

NACK, Frank, HARDMAN, Lynda. **Denotative and connotative semantics in hypermedia**: proposal for a semiotic-aware architecture. CWI, 2002.

NENTWICH, M. *Cyberscience: research in the age of the Internet*. Vienna: Austrian Academy of Sciences Press , 2003.

NIRENBURG, Sergei; RASKING, Victor. **Ontological semantics, formal ontology, and ambiguity**. 2001.

NOY, N.F., MCGUINNESS, D.L. **Ontology development 101**: A guide to creating your first ontology. Technical Report KSL-01-05, Stanford Knowledge Systems Laboratory. 2001.

OBJECT Management Group. **UML Resource Page**. 2006. Disponível em: <http://www.uml.org/>.

OLESHCHUK. Ontology based semantic similarity comparison of documents. In **14th International Workshop on Database and Expert Systems Applications (DEXA'03) 09 01 - 09, 2003**, Prague, Czech Republic. 2003. Disponível em: <<http://csdl.computer.org/comp/proceedings/dexa/2003/1993/00/19930735abs.htm>>.

OLIVEIRA, Edgard Costa. Towards a new authoring environment: an overview of ontology-based systems. In Proceedings of **EIPub'04**, Brasilia, 2004. Disponível em: <<http://elpub.scix.net/data/works/att/121elpub2004.content.pdf>>

\_\_\_\_\_. **Research Proposal: ontology-based authoring environment**. AI Term Projects, Supervisão: Frank van Harmelen, Vrije Universiteit, Amsterdam, 2004. Disponível em: <<http://www.cs.vu.nl/ai/termprojects/kr.php#authoring>>

\_\_\_\_\_. HARMELEN, Frank van; LIMA-MARQUES, Mamede. A framework for ontology-based authoring environments. In **ISWC 2004 – International Semantic Web Conference**, Hiroshima, Japão. 2004.

\_\_\_\_\_. LIMA-MARQUES, Mamede. An architecture of authoring environments for the Semantic Web. In Proceedings of **EIPub'06**, Bansko, Bulgária, 2006. Disponível em: <<http://elpub.scix.net/data/works/att/274elpub2006.content.pdf>>

ONTOWEB. Ontoweb deliverable 1.3: **A survey on ontology tools**. Technical report, IST OntoWeb Thematic Network, May 2002.

\_\_\_\_\_. Ontoweb deliverable D2.1.1 **Survey of scalability techniques for reasoning with ontologies**. August 2004.

PANG, Alex Soojung-Kim. **The work of the encyclopedia in the age of electronic reproduction. & Hypertext, the next generation: a review and research agenda**. First Monday Disponível em: <[http://www.firstmonday.dk/issues/issue3\\_9/pang/](http://www.firstmonday.dk/issues/issue3_9/pang/) , 1998>.

PIAGET, Jean. O nascimento da inteligência na criança. 4. ed. Rio de Janeiro: Zahar, 1982. 389 p.

PIETRIGA, E.: **IsaViz: A visual authoring tool for RDF**. 2002. <http://www.w3org/2001/11/IsaViz>

POPOV, B. et al. **KIM – a semantic platform for information extraction and retrieval**. Journal of Natural Language Engineering, Vol. 10, Issue 3-4, Sep 2004, pp. 375-392, Cambridge University Press.

QUAN, Dennis, HUYNH, David, KARGER, David R. Haystack: a platform for authoring end user Semantic Web applications. In **WWW' 2003**. Hungary: ACM, 2003.

QUINT, Vincent, VATTON, I. **Towards active Web clients**. DocEng 2005. Disponível em: <<http://wam.inrialpes.fr/publications/2005/DocEng05-Quint.html>>.

SAUVÉ, Jacques. **Diagrama de atividades da UML**. Acessado em 2006. Disponível em: <[http://www.dsc.ufcg.edu.br/~jacques/cursos/map/html/uml/diagramas/atividades/diag\\_atividades.htm](http://www.dsc.ufcg.edu.br/~jacques/cursos/map/html/uml/diagramas/atividades/diag_atividades.htm)>.

SAWYER, Pete; KOTONYA, Gerald. **Software Requirements**. 2001. Disponível em: <[http://info.comp.lancs.ac.uk/publications/Publication\\_Documents/2004-Sawyer-Software%20Requirements.pdf](http://info.comp.lancs.ac.uk/publications/Publication_Documents/2004-Sawyer-Software%20Requirements.pdf)>.

SCALET, Danilo. **ISO/IEC 9126 and 14598 integration aspects: A Brazilian viewpoint**. The Second World Congress on Software Quality, Yokohama, Japan, 2000.

SCHIEN, Maarten van. **Using Semantic Web technologies for electronic publishing**. Dissertação de mestrado, 68p. Vrije Universiteit, Amsterdam, 2003.

SCHREIBER, A. et al. **Knowledge engineering and management: the CommonKADS Methodology**. The MIT Press.1999.

SEMANTIC. **Semantic Web Annotation and Authoring Tools**. 2005. Disponível em <http://annotation.semanticweb.org/tools>.

SENSO, José, Piñero, A. R. *El concepto de metadato. Algo más que descripción de recursos electrónicos*. **Ciência da Informação**, Brasília, v.32, n.2, p.95-106, maio-ago 2003.

SHARPLES, M and Perberton, L. Starting from the writer: guidelines for the design of user-centred document processors. **Computer Assisted Language Learning 2**, 37-57. 1990.

\_\_\_\_\_, GOODLET, J. and PEMBERTON, L. The Writer's Assistant: Developing a Writer's Assistant, in J. Hartley (ed, **Technology and Writing: readings in the psychology of Written Communication** (pp.209-20). London: Jessica Kingsley. 1992.

\_\_\_\_\_, van der GEEST, Thea. **The new writing environment: writers at work in a world of technology**. London: Springer-Verlag, 1996.

SHNEIDERMAN, Ben. Reflections on Authoring, Editing, and Managing Hypertext. In Edward Barret (ed.) **The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information**. MIT Press, Cambridge Mass. 1989.

SHUM, Simon B. **User's Guide to SchoolOnto Claimaker: a system for modelling and visualizing claims and argumentation in research literatures**. 2003. Disponível em: <http://claimaker.open.ac.uk/>.

SILVA, Lúcio B. **Ambigüidades da língua portuguesa: recorte classificatório para a elaboração de um modelo ontológico**. Dissertação de Mestrado, Universidade de Brasília. 2006

SMITH, J.B, LANSMAN, M. The writing environment: a cognitive basis for a computer writing environment, in B.K. Britton and S.M. Glynn (eds), **Computer writing environments: theory, research and design** (pp 17-56). Hillsdale, NJ: Lawrence Erlbaum. 1989.

SOBIESIAK, Rick. Towards a New Generation of Authoring Tools. **Proceedings of the 13th annual international conference on Systems documentation: emerging from chaos: solutions for the growing complexity of our jobs**. Savannah, Georgia, p. 157-161. ACM Press New York, NY, USA. 1996.

SOUZA, Renato Rocha, ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, v. 33, n.1, p. 132-141, jan-abril 2004. Disponível em: <http://www.ibict.br/cionline/viewarticle.php?id=71>.

STUCKENSCHMIDT, Heiner. **Ontology-based information sharing in weakly-structured environments**. PhD thesis. Vrije Universiteit Amsterdam. 2003.

\_\_\_\_\_, et al. **Generating and managing metadata for Web-based information systems**. 2003.

SUMNER, Tamara; SHUM, Simon B. From documents to discourse: shifting conceptions of scholarly publishing. **CHI'98: Human Factors in Computing Systems. Proceedings**. ACM Press: New York, 1998. Disponível em: <http://kmi.open.ac.uk/kmi-abstracts/kmi-tr-50-abstract.html>.

SURE, York, et al. **Methodology for development and implementation of ontology based knowledge management applications**. ACM Press, Sigmod Volume 31, Issue 4 New York, NY, USA , 2002. Disponível em: <http://portal.acm.org/citation.cfm?id=637414>.

TALLIS, Marcelo. Semantic Word processing for content author, in **K-CAP 2003, Second International Conference on Knowledge Capture 2003**, Florida, USA. Disponível em: <http://mr.teknowledge.com/DAML/>.

\_\_\_\_\_. KIM, Jihie and GIL, Yolanda. **User studies of knowledge acquisition tools**. Disponible em: <<http://sern.ucalgary.ca/KSI/KAW/KAW99/papers/Tallis1/kaw99-6-7.pdf>>.

TOL, Martin van de. The abstract as an orientation tool in modular electronic articles. In proceedings: **First International Conference on Document Design**, Tilburg, 1998.

UREN, Victoria, SHUM, Simon B, et al. Writing the Web: scholarly publishing and argument in hyperspace, in **WWW'03** Proceedings of the 12th international conference on World Wide Web, Hungary, 2003.

USCHOLD, M., JASPER, R. A framework for understanding and classifying ontology applications. Proceedings of **IJCAI-99**, Sweeden, 1999. Disponible em: <<http://citeseer.ist.psu.edu/257410.html>>

VELARDI, Paola. **Using text processing techniques to automatically enrich domain ontologies**. ACM Press, New York, 2001.

VELTMAN, Kim. Frontiers in Electronic Media, *Interactions*, vol. 4, Isss.4, pp.32, 1997.

W3C. **XML schema datatypes in RDF and OWL**. W3C Working Draft 27 April 2005. Disponible em: <<http://www.w3.org/TR/2005/WD-swbp-xsch-datatypes-20050427>>.

\_\_\_\_\_. **Authoring challenges for device independence**. W3C Working Group Note 1 September 2003. Rhys Lewis (Editor). Disponible em: <<http://www.w3.org/TR/acdi/>>.

\_\_\_\_\_. **Authoring techniques for XHTML & HTML internationalization**: Specifying the language of content 1.0. W3C Working Draft 24 February 2005. Richard Ishida (editor). Disponible em: <<http://www.w3.org/TR/2005/WD-i18n-html-tech-lang-20050224/>>.

\_\_\_\_\_. **Authoring tool accessibility guidelines 2.0**. Working draft 23 November, 2005. J. Treviranus, J. Richards and M. May, 3eio4w. 2005. Disponible em: <<http://www.w3.org/TR/2005/WD-ATAG20-20051123/>>.

\_\_\_\_\_. **Compound document use cases and requirements** Version 2.0. W3C Working Draft 19 December 2005. Steve Speicher and Petri Vuorimaa (editors). Disponible em: <<http://www.w3.org/TR/CDFReqs/>>.

\_\_\_\_\_. **Device independent authoring language (DIAL)**. W3C Working Draft 16 May 2006. Kevin Smith (Editor). Disponible em: <<http://www.w3.org/TR/2006/WD-dial-20060516/>>.

\_\_\_\_\_. **Extensible markup language XML 1.0**. Recommendation. February 2004. Disponible em: <<http://www.w3.org/TR/REC-xml/>>.

\_\_\_\_\_. **Glossary of terms for device independence**. W3C Working Draft 18 January 2005. Rhys Lewis (author) Disponible em: <<http://www.w3.org/TR/di-gloss/>>.

\_\_\_\_\_. **HTML Tidy**. 2006. <http://www.archivesat.com/html-tidy.w3.org/> e <http://www.w3.org/People/Raggett/tidy/> .

\_\_\_\_\_. **HyperText Markup Language (HTML) home page**. 2006. Disponible em: <<http://www.w3.org/MarkUp/> e <http://validator.w3.org/>>.

\_\_\_\_\_. **Ontology driven architectures and potential uses of the Semantic Web in systems and software engineering**. Editors' Draft, 2006. Disponible em: <<http://www.w3.org/2001/sw/BestPractices/SE/ODA/060211/>>.

\_\_\_\_\_. **OWL Web Ontology Language: use cases and requirements**. W3C Recommendation 10 February 2004. Jeff Heflin, editor. Disponible em: <<http://www.w3.org/TR/webont-req/>>.

\_\_\_\_\_. **OWL, Web Ontology Language overview**. 2004. D. McGuinness e F. van Harmelen (editor). Disponível em: <<http://www.w3.org/TR/2004/REC-owl-features-20040210/>>.

\_\_\_\_\_. **Quick guide to publishing a thesaurus on the Semantic Web**. W3C Working Draft 17 May 2005. Alistair Miles (editor). Disponível em: <<http://www.w3.org/TR/2005/WD-swbp-thesaurus-pubguide-20050517/>>.

\_\_\_\_\_. **RDF/XML syntax specification (Revised)**. W3C Recommendation 10 February 2004. Dave Baggett. Editor. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>>.

\_\_\_\_\_. **SVG tiny version 1.2 requirements**. W3C Working Draft 09 December 2003. Chris Lilley (editor). Disponível em: <<http://www.w3.org/TR/SVGTiny12Reqs>>.

\_\_\_\_\_. **The essentials of a specification**. Tim Berners-Lee (editor), 1999. Disponível em: <<http://www.w3.org/1999/09/specification.html>>.

\_\_\_\_\_. **The recommendation XHTML™ 1.0 The Extensible HyperText Markup Language**. Disponível em: <<http://www.w3.org/TR/xhtml1/>>.

\_\_\_\_\_. **The W3C hypertext markup page**. 2005. Disponível em: <<http://www.w3.org/MarkUp/> e <<http://www.w3.org/XML/>>.

\_\_\_\_\_. **The W3C Semantic Web Best Practices & Deployment Working Group**. 2006. Disponível em: <<http://www.w3.org/2001/sw/BestPractices/>>.

\_\_\_\_\_. **W3C technical reports and publications**. Disponível em: <<http://www.w3.org/TR/#Recommendations>>. Acessado em junho de 2006.

\_\_\_\_\_. **WCAG20 Web content accessibility guidelines 2.0 (Working Draft)**, 2006. W. Chisholm, G. Vanderheiden, and J. White (Editores). Disponível em: <<http://www.w3.org/TR/WCAG20/>>.

\_\_\_\_\_. **XHTML™ 1.0 The Extensible HyperText Markup Language (Second Edition)**. A Reformulation of HTML 4 in XML 1.0. W3C Recommendation 26 January 2000, revised 1 August 2002. Steven Pemberton et al (Editores). Disponível em: <<http://www.w3.org/TR/xhtml1/#acks>>.

\_\_\_\_\_. **XHTML™ 2.0**. W3C Working Draft 27 May 2005. Jonny Axelsson et al (editores). Disponível em: <<http://www.w3.org/TR/2005/WD-xhtml2-20050527/>>.

\_\_\_\_\_. **XHTML™ document profile requirements**. Document profiles - a basis for interoperability guarantees. W3C Working Draft 6th September 1999. Dave Raggett, Peter Stark, Ted Wugofski (Editors). Disponível em: <<http://www.w3.org/TR/xhtml1-prof-req/>>.

\_\_\_\_\_. **XML Processing Model Requirements and Use Cases**. W3C Working Draft 11 April 2006. Alex Milowski (editor). <<http://www.w3.org/TR/2006/WD-xproc-requirements-20060411/>>.

\_\_\_\_\_. **XML Schema Requirements**. W3C Note 15 February 1999. Ashok Malhotra e Murray Maloney (Editores). Disponível em: <<http://www.w3.org/TR/NOTE-xml-schema-req>>.

WAARD, Anita de. **From documents to knowledge bases**: a proposal for scientific authoring in a linked environment. 2003.

\_\_\_\_\_. **Science publishing and the Semantic Web, or: why are you reading this on paper?** ECSW, 2005. Disponível em: <[www.cs.uu.nl/people/anita/papers/deWaardECSW2005.pdf](http://www.cs.uu.nl/people/anita/papers/deWaardECSW2005.pdf)>.

\_\_\_\_\_. KIRCZ, J. Metadata in science publishing. In Proceedings **Conferentie Informatiewetenschap Technische Universiteit Eindhoven**, 2003. Disponível em: <<http://www.kra.nl/Website/Artikelen/Metadata-wi-2003.htm>>

WALKER, Janet H. **Authoring tools for complex document sets**. In Edward Barret (ed.) *The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information*. MIT Press, Cambridge Mass. 1989.

WIKIPÉDIA. **Caso de uso**. Disponível em: <[http://pt.wikipedia.org/wiki/Caso\\_de\\_uso](http://pt.wikipedia.org/wiki/Caso_de_uso)>. Acessado em 2006.

\_\_\_\_\_. **Diagrama de caso de uso**. Disponível em: <[http://pt.wikipedia.org/wiki/Diagrama\\_de\\_Caso\\_de\\_Uso](http://pt.wikipedia.org/wiki/Diagrama_de_Caso_de_Uso)>. Acessado em 2006.

ZELKOWITZ, M.V. & Wallace, D. **Experimental models for validating computer technology**. IEEE Computer. 1998. Disponível em: <<http://www.cs.umd.edu/~mvz/pub/computer97.pdf>> .

ZENG, Marcia. **Science learning spaces**. The 6th Networked Knowledge Organization Systems/Sources (NKOS) Workshop. May 31, 2003 in Houston, TX as part of JCDL 2003. In: *Building a More Meaningful Web: From Traditional Knowledge. Organization Systems to New Semantic Tools*, Dagobert Soergel (Org.) University of Maryland.

## ANEXO I

**Tabela 3 : Compilação dos critérios para análise de requisitos**

<b>Autor</b>	<b>Título do Trabalho</b>	<b>Objetivo</b>	<b>Crítérios Nível I</b>	<b>Crítérios Nível II</b>
Handschuh&Staab (2002)	Design Rational - Linking Requirements with CREAM Modules	Authoring and annotation of Web pages in CREAM	Requirements: Document editor Ontology guidance Crawler Annotation Inference Server Document Management Metadata recognition General Information Extraction Meta Ontology Content Generation General Problem Consistency Proper Reference Avoid Redundancy Relation Metadata Maintenance Ease of use Efficiency	
Gruber (1993)	Towards principles for the design of ontologies used for knowledge sharing.	Evaluate the products of ontology engineering	Clarity Coherence Extendibility Minimal Encoding Bias Minimal Ontological Commitment	
Gómez-Pérez (1995)	Criteria do verify knowledge sharing technology	Define a set of criteria to use to score efficiency of KST	Verification of the structure/architecture  Verification of the lexicon and syntax  Verification of the content	Soundness  Correctness  Consistency, Completeness, Conciseness, Expandability and Sensitiveness
Uschold&Jasper (1999)	A framework for understanding and classifying ontology applications	Identify categories of ontology applications and scenarios, purpose, role of ontology, supporting technologies and actors	Neutral Authoring  Common Access to Information  Indexing for Search	
Ontoweb (2004)	A survey on ontology Tools		Ontology-building tools  Ontology-based annotation tools  Ontology merge and integration tools  Ontology evaluation tools  Ontology storage and querying tools	-General Description -Software architecture and tool evolution -Interoperability with other tools and languages -Knowledge representation expressivity -Inference services -Usability -Documentation -Tutorial Material -Available modes of working -Automation -Interoperability -Ontology related points -Kind of documents that can be annotated

<b>Autor</b>	<b>Título do Trabalho</b>	<b>Objetivo</b>	<b>Crítérios Nível I</b>	<b>Crítérios Nível II</b>
OntoWeb (2002)	Survey on ontology tools		Screen snapshot URL References Documentation Tutorial Material Available modes of working Automation Interoperability with other ontology development tools Ontology related points  Kind of document that can be annotated Usability aspects	a. where ontologies can be retrieved from b. where populated ontologies can be written to c. the underlying languages(s) for the ontology definition and for the ontology notations d. whether browsing of concept/properties/relations is possible e. whether the tool provides restricted values f. whether the tool checks constraints  a. how easy is to learn the system? b. How easy is it in everyday use? c. How efficient is the tool?
Angele et al (2003)	EFFORT	Evaluation framework for ontologies and related technologies	Ontology Properties  Technology Properties	Language conformity (syntax) Consistency (semantics) Interoperability (e.g. Semantics) Turn around ability  Performance Memory allocation  Scalability Integration into frameworks Connectors and interfaces
Denny (2002)	Ontology building: a survey of editing tools	Identify in the analyzed tools the following aspects	Tool name  Version Release Date Source Modeling Features/Limitations Base Language Web support & Use Import/Export formats Graph view Consistency checks Multi-user support Merging Lexical support Information extraction Comments More information Contact	

<i>Autor</i>	<i>Título do Trabalho</i>	<i>Objetivo</i>	<i>Crítérios Nível I</i>	<i>Crítérios Nível II</i>
Denny (2004)	Ontology Tool Survey, Revisited.		<p>Abstraction for knowledge modeling</p> <p>Visual/intuitive navigation of ontology</p> <p>Reasoning and problem solving facilities</p> <p>Ontology alignment and data resource integration</p> <p>Support of standard industry domain and core vocabularies</p> <p>Natural language processing</p> <p>Versioning control</p> <p>Built-ins (wizards) for best practice methods</p> <p>Information extraction facilities</p> <p>Features to learn user's editing style and needs</p> <p>Collaborative development support</p> <p>Ontology support for contexts</p>	
ISO/IEC 9126-1:2001	<p>Software engineering - product quality</p> <p>Part 1: quality model</p> <p>Part 2: external metrics</p> <p>Part 3: internal metrics</p> <p>Part 4: quality in use metrics</p>	Quality evaluation of software	<p>Functionality</p> <p>Reliability</p> <p>Usability</p> <p>Efficiency</p> <p>Maintainability</p> <p>Portability</p>	<p>suitability</p> <p>accuracy</p> <p>interoperability</p> <p>security</p> <p>maturity</p> <p>fault tolerance</p> <p>recoverability</p> <p>understandability</p> <p>learnability</p> <p>operability</p> <p>attractiveness</p> <p>time behavior</p> <p>resource utilization</p> <p>analyzability</p> <p>changeability</p> <p>stability</p> <p>testability</p> <p>adaptability</p> <p>installability</p> <p>co-existence</p> <p>replaceability</p>
MacKnight & Balagopalan (1989)	Evaluation Tool for measuring authoring system performance	Score systems tasks by the 3 dimensions.	<p>Functionality</p> <p>Flexibility</p> <p>Productivity</p>	<p>Authoring Environment</p> <p>Text Creation</p> <p>Text Editing</p> <p>Graphics Creation</p> <p>Graphics Editing</p> <p>Animation</p> <p>Interactive Video</p> <p>Interactive Sound</p> <p>Artificial intelligence</p> <p>Instructional Strategy</p> <p>Calculations</p> <p>Student Management</p>

<b>Autor</b>	<b>Título do Trabalho</b>	<b>Objetivo</b>	<b>Critérios Nível I</b>	<b>Critérios Nível II</b>
Cristea et al (2004)	Evaluation criteria for and adaptive hypermedia authoring tool	Defined a set of requirements to be tested and established some evaluation goals to be achieved	Goal Point of View  User Interface Point of View	Collaboration  Completeness Adaptivity Design range Ease of use Robustness Complexity of interface
Oliveira (2004)	Requirements for Authoring Environments	Features to be present in Semantic Web Authoring environments	Authoring Semantic Mark-up and Annotation Semantic Browsing Metadata generation Multi-ontology support Ontology visualization Thesaurus-based term/concept choice Information Extraction Text Engineering Resource discovery environment Argumentation and peer review Publishing	

## **Anexo II - Especificação de Requisitos do Sistema DocOnto**

### **A. Introdução**

Esta é a especificação de requisitos do ambiente de autoria baseado em ontologia para a Web Semântica, pela qual apresentaremos os itens necessários para se estabelecer os propósitos que a solução proposta busca alcançar, o escopo em que se dá a aplicabilidade do ambiente de autoria, as definições e referências utilizadas para elaborar esta especificação, assim como uma visão geral da solução. Ela foi elaborada com base na norma IEEE Std 830-1998.

#### **A.1 Objetivos da especificação de requisitos**

A presente especificação de requisitos tem os seguintes objetivos principais:

- i) explicitar os requisitos necessários para a composição de um ambiente de autoria para a Web Semântica, baseado em ontologias.
- ii) organizar as características do sistema, com base no estado-da-arte, com base em pesquisa específica de requisitos, assim como em novos requisitos propostos, decorrentes da criação do novo modelo de ambiente ora apresentado.
- iii) permitir a validação do modelo proposto de arquitetura de ambientes, por meio de uma lista completa de atores do sistema, funcionalidades, tecnologias, desempenho e comunicação.
- iv) orientar a elaboração das demais partes do anteprojeto de desenvolvimento de software, as quais foram desenhadas e apresentadas nos próximos itens deste trabalho: os diagramas de caso de uso do sistema; o diagrama de domínio do sistema; o diagrama de atividades e os diagramas de seqüência de atividades.
- v) associar os requisitos do sistema e características percebidas e propostas, com as referências ou justificativas apresentadas, para se obter uma correta explicitação da idéia proposta e sua conseqüente compreensão e validação pelas partes envolvidas.

As partes envolvidas na análise e validação da especificação de requisitos ora proposta, ou seja, o público-alvo deste documento são os seguintes.

1. Analistas de sistemas – grupo de pessoas responsáveis pela análise dos requisitos e desenho dos diagramas de forma padronizada e adequada às exigências da prática de desenvolvimento de software. Os diagramas representam requisitos parciais para a elaboração do anteprojeto de desenvolvimento de software e que juntos compõem a proposta de ambiente do autoria baseado em ontologias para a Web Semântica.
2. Banca examinadora da tese de doutorado – grupo de doutores responsáveis pelo julgamento da solução proposta para o reconhecimento da sua aplicabilidade e viabilidade no âmbito da Ciência da Informação.

3. Comunidade acadêmica – pesquisadores e professores envolvidos na pesquisa de soluções para autoria de documentos, interessados por propostas de soluções de desenvolvimento de ambientes de software para a Web Semântica.

4. Desenvolvedores de software – grupo de programadores envolvidos com o desenvolvimento da solução em termos de código de programação lógica de softwares.

5. Fornecedores – grupo de pessoas que detêm tecnologias específicas referenciadas nesta especificação e que cumprem a função de integradores de suas soluções com aquela proposta neste trabalho, buscando validar a viabilidade de implantação da solução em conjunto com as tecnologias existentes.

5. Agências financiadoras de pesquisa – entidades que oferecem recursos de investimento para a pesquisa científica, e que estejam interessadas em apoiar a pesquisa no âmbito da Web Semântica, visando à compreensão dessas agências quanto aos requisitos necessários para o desenvolvimento da solução proposta e a viabilidade de execução com base em custos e tempo necessário de desenvolvimento para implementação.

## **A.2 Escopo**

Nesta seção é apresentado o sistema que se busca especificar, por meio de explicações sobre o que ele irá fazer; descrições de como o sistema está sendo especificado, incluindo seus benefícios e objetivos.

O sistema ora especificado é destinado ao uso em qualquer ambiente de produção de conhecimento, especialmente desenhado para a autoria de documentos com representação do conhecimento a partir de ontologias. O objetivo principal do sistema é de prover aos autores uma infra-estrutura ontológica para que possam localizar conceitos, interligar os conceitos com fontes de informação, representar os metadados do documento de forma automática ou semi-automática e permitir a navegação ontológica e incrementar a recuperação da informação em meio eletrônico, durante a autoria do documento.

## **A.3 Definições, acrônimos e abreviações**

Nesta seção são apresentadas as definições de termos utilizados nesta especificação, para orientar os leitores.

<b>Ambiente de descoberta (<i>discovery</i>)</b> – ambiente eletrônico no qual se dá a pesquisa da informação.
--

<b>Ambiente de ontologias</b> – ambiente eletrônico no qual se dá a representação semântica da informação a partir de relações lógicas e contextuais.
---

<b>Ambiente de produção</b> – ambiente eletrônico no qual se dá a autoria de documentos.
--

<b>Ambigüidade</b> – expressão da língua que possui vários significados distintos, podendo ser compreendida de diferentes maneiras pelo receptor.
---

<b>Análise léxica</b> – atividade realizada pelos extratores de informação para fins de identificação dos elementos que constituem um texto no seu sentido sintático, como artigo, sujeito, verbo, objeto, etc.
<b>Autor</b> – pessoa ou grupo que redige um documento e pode deter a sua autoria.
<b>Autoria</b> – atividade de registro da informação em ambiente eletrônico, a qual faz uso de tecnologias correntes de produção de textos, documentos e seus correspondentes imagéticos.
<b>Autoria em ambiente aberto (Open-world authoring)<sup>2</sup></b> - situação na qual as escolhas feitas durante a autoria ficam registradas em uma base de conhecimento, para que a informação implicitamente codificada em um documento torne-se explícita na base de conhecimento e pode ser reutilizada para simplificar a autoria de novos documentos.
<b>Autoria em ambiente fechado (Closed world authoring)<sup>3</sup></b> - situação de produção de documento na qual uma base de conhecimento é utilizada para restringir o uso de documentos possíveis e orientar as opções de associação semântica realizadas pelos usuários.
<b>Documento composto</b> – documento da Web que pode conter simultaneamente diversos elementos de informação como texto, imagem, <i>links</i> etc., ou também apresentar diferentes padrões de formato e linguagem.
<b>DTD</b> – Document Type Declaration – elemento que identifica o tipo de documento que está sendo criado, para fins de classificação e representação do mesmo.
<b>Engenheiro de ontologia</b> – pessoa responsável por verificar a consistência da ontologia do documento criada pelo autor, a associação dessa com as ontologias existentes, a criação de novas relações ontológicas sugeridas pelo autor.
<b>Inconsistência</b> – a presença de algum erro ou ausência de relação entre conceitos de uma instância na representação ontológica.
<b>Indexador semântico</b> – mecanismo eletrônico que realiza a associação de fontes de informação do ambiente de descoberta com as instâncias reconhecidas na ontologia.
<b>Instância</b> – unidade de uma ontologia utilizada para representar elementos específicos, referente a um termo ou conceito.
<b>Ligar</b> – unir dois objetos em um ambiente eletrônico por meio de um <i>link</i> , ou seja, uma referência identificada por uma URI. Advém do termo “ <i>to link</i> ” em inglês.
<b>Marcação ontológica</b> – destaque de termos em um documento de forma a demonstrar a existência de associação entre o termo e classes ou instâncias de uma ontologia.
<b>Metadado</b> – elemento de identificação de conteúdo de um documento, utilizado para representar o documento no tocante à estrutura, título, data, etc. para fins de recuperação da informação.

<sup>2</sup> Dimetman, Marc. Text authoring, knowledge acquisition and description logics. 2003

<sup>3</sup> Idem.

<b>Meta-ontologia</b> – é uma ontologia de alto nível que geralmente apresenta os principais componentes conceituais de outras ontologias de domínio.
<b>Movimentação da ontologia</b> – habilidade em se visualizar a ontologia a partir de qualquer uma de suas classes e instâncias para facilitar a navegação ontológica.
<b>Navegação ontológica</b> – uso de ontologias para realizar navegação na Web a partir de relações semânticas existentes entre os conceitos e termos nelas existentes.
<b>Ontobroker</b> – pessoa ou empresa que realiza trabalhos de elaboração ou hospedagem de ontologias de domínios.
<b>Ontologia de domínio</b> – sistemas de conceitos inter-relacionados e agrupados conforme as características de áreas de conhecimento específicas.
<b>Ontologia global</b> – sistemas de conceitos genéricos e que podem ser usados para organizar as ontologias de domínio.
<b>Parser/Extrator</b> – mecanismo eletrônico que realiza a categorização de entidades léxicas a partir de uma análise automática.
<b>Parseamento</b> – atividade realizada pelo parser para identificar entidades léxicas em uma frase.
<b>Produção de documentos</b> – é o ciclo de vida de um documento no que se refere a criação, digitação, edição, representação, revisão e publicação de um documento.
<b>Raciocínio ontológico</b> – mecanismo de reconhecimento automático realizado por um agente inteligente que associa um termo presente em um documento a outros equivalentes, presentes em ontologias com suas respectivas instâncias e relações semânticas.
<b>Sistema DocOnto</b> – nome da solução proposta do ambiente de autoria baseado em ontologias para a Web Semântica, ora especificado.
<b>Termo</b> – unidade de representação contida em classes ou instâncias de uma ontologia, à qual se atribuem relações semânticas com os demais elementos da ontologia ou de outras ontologias.
<b>Tesouro</b> – estrutura de representação da informação que associa termos aos seus conceitos correlatos por relações de hierarquia e preferência de uso.
<b>Validação</b> – ato de revisar e aprovar novas relações semânticas feitas em uma ontologia por um autor de documentos.
<b>Visualização de ontologia</b> – mecanismo de representação gráfica da informação contida nas ontologias, por meio de modelos gráficos ou textuais.

## **A.4 Visão geral da especificação de requisitos**

A especificação de requisitos do ambiente de autoria está organizada em 2 partes principais:

1. Descrição Geral do Ambiente de Autoria – nessa seção é descrita a perspectiva do ambiente em termos de componentes funcionais, as interfaces do sistema: software e hardware, as funções do sistema, as características dos usuários, as limitações e demais restrições que o ambiente apresenta em sua fase atual.

2. Especificação de Requisitos do Ambiente – nessa seção são apresentados os requisitos do ambiente de autoria, levando-se em conta a composição dos módulos do sistema. Esta seção foi dividida em duas grandes subseções: Configuração e Uso do ambiente. Respectivamente, foram listados os requisitos de cada subseção, considerando as funcionalidades dos ambientes sugeridos: ambiente de autoria, ambiente de extração da informação, ambiente ontológico e ambiente de descoberta.

## **B. Perspectiva do Ambiente**

### **B.1 Interfaces do Sistema**

Neste item são descritas cada uma das interfaces do sistema, identificando-se a funcionalidade do software para satisfazer o requisito do sistema.

O sistema é composto por 4 interfaces:

1. editor de texto/autoria, onde é feita a entrada de dados para produção de documento;
2. visualizador de ontologias, onde é apresentada a ontologia utilizada como referência semântica para a associação com os termos digitados;
3. editor de ontologias, onde é oferecido ambiente de engenharia ontológica para a inclusão de novos conceitos ou relações com instâncias da ontologia;
4. navegador pelo qual é feita a navegação semântica em fontes de conhecimento relacionadas, indicadas por meio das ontologias.

### **B.2 Interface do Usuário**

A interface do usuário é composta por três módulos:

- tela do ambiente de edição de textos;
- tela do ambiente de ontologias (visualização e edição);

- tela do ambiente de descoberta (navegação).

Nessa interface, apresentada a seguir, buscamos ilustrar como os três módulos podem ser integrados e disponibilizados ao autor do documento. À esquerda, enquanto digita o texto, os termos reconhecidos na ontologia são anotados (em cores diferentes) e associados à respectiva instância na ontologia (à direita, superior). As fontes de informação referentes aos conceitos digitados aparecem listadas na tela de baixo (à direita).

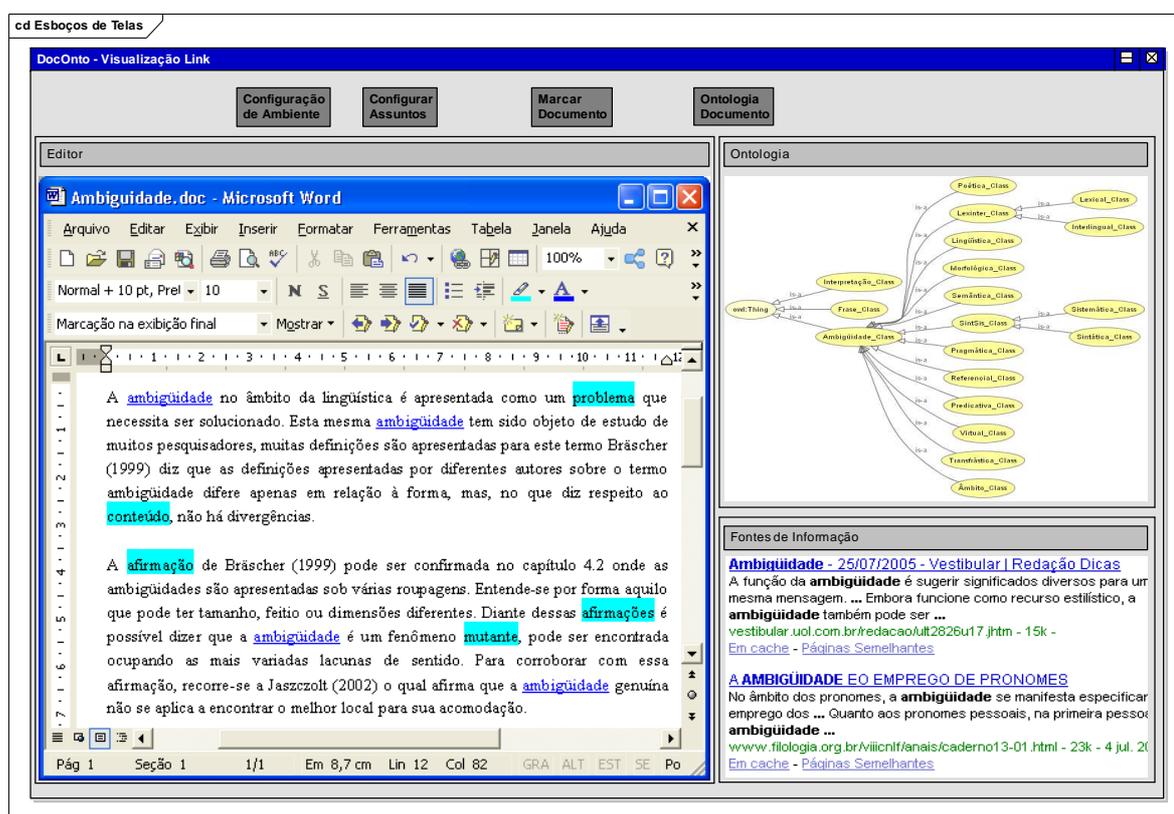


Figura 19 – Tela do ambiente de autoria para a Web Semântica

### B.3 Interface de software

Nesta seção são especificados o uso de outros softwares e interfaces com outras aplicações.

- Software de edição de textos e de autoria
- Software de extração da informação (parser)
- Software de edição de ontologias
- Software de visualização de ontologias
- Software de anotação e marcação semântica
- Software de navegação na Web.

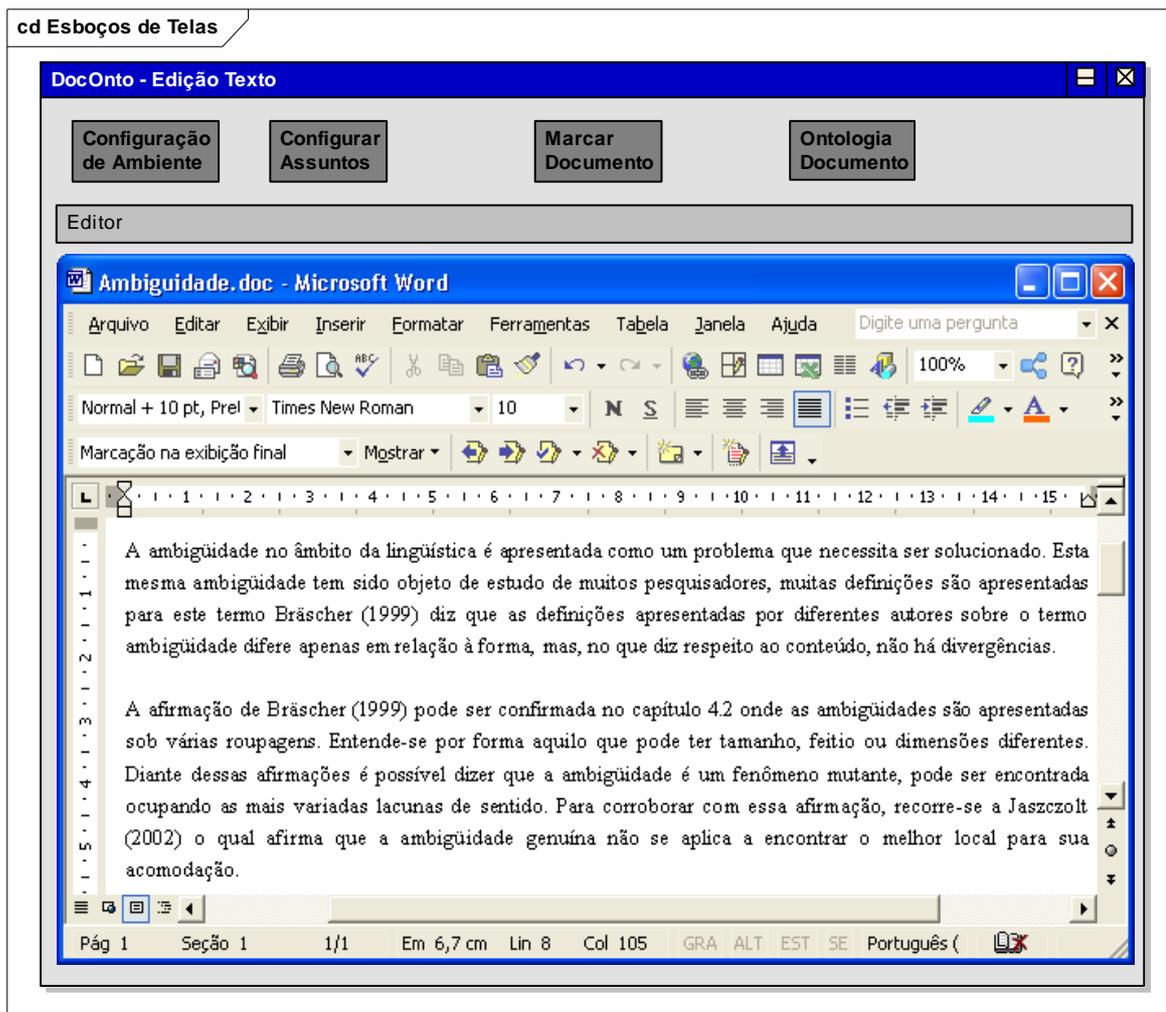


Figura 20 – Tela do editor do ambiente de autoria

Nesta tela é apresentada a interface do sistema em termos gráficos. O autor terá acesso primeiramente a uma tela de editor de textos comum. Neste caso, utilizamos o MS Word para ilustrar um editor de textos amplamente utilizado, familiar a usuários e com funcionalidades específicas do software.

O autor tem a liberdade de digitar o texto livremente e fazer as inferências ontológicas antes, durante ou depois da autoria. Se desejar, pode visualizar em tela os três principais módulos do ambiente de autoria baseado em ontologias.

### **C. Funções do produto especificado**

Nesse item é apresentado um resumo das principais funções que o sistema irá desempenhar. As funções são organizadas de forma a permitir que os requisitos estejam estruturados e listados para facilitar a compreensão das características do mesmo por parte do público-alvo. A norma IEEE Std 830-1998 recomenda que sejam utilizados métodos de descrição tanto textuais quanto gráficos, de forma a demonstrar as diferentes funções e suas relações, de uma forma bastante sucinta e superficial. Mais adiante, no item H desta especificação, essas funcionalidades encontram-se em um nível de detalhamento maior, com suas respectivas representações por meio dos diagramas constantes do anteprojeto.

- Suporte à interpretação do documento em produção, por meio da associação de uma camada de ontologias e permitindo a anotação semântica e navegação conceitual.
- Um mecanismo de mediação entre um documento não estruturado e estruturas semânticas existentes em ontologias disponíveis e relacionadas ao assunto em questão.
- Ontologias utilizadas para marcação podem ser atribuídas pelo autor, para autoria em ambiente aberto, ou automaticamente associadas, para autoria em ambiente fechado.
- Um serviço de raciocínio ontológico para representar termos e conceitos de documentos e suas relações com um ambiente interligado via ontologias.
- Interferir na capacidade do autor em criar novos termos ou usar termos variantes para designar conceitos, com o suporte de ontologias para restringir o uso de termos em textos quando da autoria em ambiente fechado.
- Oferecer a opção de associação a uma ou mais ontologias, em assuntos e idiomas específicos.

#### ***D. Características dos usuários do sistema***

O ambiente de autoria baseado em ontologia para a Web Semântica destina-se ao uso de autores de documentos eletrônicos em geral, ou seja, qualquer pessoa que se utiliza de ambientes informatizados para a elaboração de um texto e suas respectivas atividades de edição e publicação, levando em consideração as potencialidades que a Web Semântica oferece.

Os autores podem ser pessoas que produzem literatura primária, secundária ou terciária, independentemente do propósito da documentação e do perfil dos usuários, sejam eles acadêmicos, especialistas, pesquisadores de pós-graduação, professores, escritores, jornalistas, literatos etc.

Os revisores e críticos compõem um segundo grupo de usuários beneficiados pelo sistema, pois podem dispor do documento para revisão, interligado a um ambiente ontologicamente representado para que se tenha acesso aos conceitos, idéias, referências e citações utilizadas nos textos do documento.

Os engenheiros de ontologias formam um terceiro grupo de usuários do ambiente, também beneficiados pela representação ontológica do texto. Eles têm duas funções básicas: i) dar suporte técnico na construção de novas relações ontológicas a usuários menos experientes com a engenharia ontológica; ii) validar a consistência da ontologia gerada por usuários experientes que construíram suas próprias ontologias ou editaram as existentes, por meio de um processo de revisão dos conceitos e das relações criadas.

Os leitores do documento, ou seja, os usuários do documento gerado, são beneficiados pelo ambiente no sentido inverso: ao acessarem o documento via ambiente de autoria baseada em ontologia, poderão também ter acesso à ontologia

do documento, às ontologias associadas ao documento assim como às bases de conhecimento equivalentes. Os leitores poderão ter acesso ao documento diretamente por meio das palavras-chave, das ontologias, facilitando a localização de assuntos dentro do documento, assim como a localização de documentos correlatos dentro da estrutura semântica apresentada.

### ***E. Limitações do ambiente***

A principal limitação atual do ambiente é quanto a ausência de ontologias em todos os domínios possíveis, restringindo a autoria de documentos sobre assuntos cujas ontologias já estejam disponíveis e com profundidade de abordagem suficiente para a representação detalhada do conteúdo existente.

A Web Semântica ainda é um ambiente em desenvolvimento, suas tecnologias estão sendo especificadas e propostas. Logo, para fins de validação experimental, seria necessária uma infra-estrutura tecnológica da Web Semântica que hoje ainda não se encontra disponível.

Uma vez que a Web Semântica ainda é uma proposta recente e as tecnologias disponíveis para a composição do ambiente são limitadas, ainda há poucas opções de solução tecnológica para tanto. Logo, deve-se levantar quais provedores de soluções estão dispostos a integrar as tecnologias para alcançar o objetivo proposto. Após esse levantamento e esforço de integração será possível identificar quais soluções são interoperáveis, assim como novas soluções que deverão ser desenvolvidas para atingir os objetivos propostos.

Um outro aspecto limitante é quanto ao conhecimento dos usuários em desempenhar as funções ora propostas, que anteriormente não eram exigidas dos autores, mas que com a Web Semântica tornam-se um novo paradigma de produção. Deve-se considerar o treinamento e orientação dos atores envolvidos em usufruir da tecnologia de forma a conduzi-los aos resultados esperados.

Um estudo aprofundado dos protocolos de comunicação, linguagens de representação, assim como os requisitos de segurança do ambiente devem ser conduzidos à luz da Web Semântica, para que o ambiente seja de fato implementado em seu aspecto tecnológico. Esse estudo também permitirá a interoperabilidade entre as soluções propostas, de forma integrada e comunicável entre si, para se atingir os propósitos do ambiente.

### ***F. Suposições e Dependências***

Nesta seção são listados os fatores que podem afetar os requisitos especificados neste documento.

O sistema proposto tem como premissa principal o uso de tecnologias disponíveis no mercado para a implementação do ambiente de autoria baseada em ontologias. Assim sendo, ao se unirem tecnologias existentes, o esforço maior de implementação está na interoperabilidade entre esses ambientes, que até o presente momento, encontram-se ainda sem interligação. Assim, supõe-se que estudos de viabilidade de integração seriam anteriores ao desenvolvimento de um novo ambiente conforme as especificações.

Um aspecto de dependência da tecnologia proposta é com relação às características que os sistemas existentes apresentam, ou seja, será necessário considerar as funções oferecidas correntemente por esses sistemas como base para a viabilidade do ambiente proposto. Como exemplo, podemos citar o caso do editor de textos: ao ser escolhido o MSWord™ como ambiente de edição de texto, todas as suas funcionalidades são adotadas como opções de edição. O mesmo ocorre com os demais ambientes, como os de ontologia e de navegação semântica. As tecnologias integrantes escolhidas determinarão muitas das funções e características de uso do ambiente em questão. No entanto, o ambiente ora especificado propõe-se a apresentar um requisito de interoperabilidade entre quaisquer soluções que apresentem requisitos de funcionalidade compatíveis com aqueles ora propostos, considerando-se a linguagem e tecnologia adotadas para a sua operabilidade.

### **G. Supressão de requisitos**

Alguns requisitos propostos podem ser implementados apenas em versões futuras do sistema. Isto se dá por diversos aspectos apresentados nos itens anteriores desta especificação. Dentre eles destacamos a associação das ontologias com o ambiente de descoberta. Devem ser estudadas formas de identificação, junto às ontologias, de fontes de informação relacionadas aos respectivos conceitos nela presentes. Esse aspecto é bastante complexo quando se leva em conta, por exemplo, a volatilidade da informação na Web, em que há significativas mudanças de identificadores de fontes de informação (URLs), além do problema da temporalidade dos documentos disponíveis.

Um requisito que foi retirado do escopo deste trabalho refere-se à revisão do documento por pares. Para esse elemento, apesar de estar claramente identificado como beneficiado por essa solução, é necessário pesquisar mais detalhadamente as possibilidades de se dispor de um ambiente de discussão semântica e de explicitação de estrutura retórica do documento, ontologicamente representada, para a revisão de pares.

As interfaces de comunicação foram suprimidas da presente especificação no estágio atual, por se tratarem de características dependentes de tecnologias.

### **H. Requisitos específicos**

Nesta seção são apresentados os requisitos de software em nível de detalhamento suficiente para que os designers (analistas de sistemas e desenvolvedores) possam definir o ambiente a partir das informações prestadas. Deve também permitir a análise dos requisitos de modo que o público-alvo deste documento possa compreender os objetivos e funcionalidades apresentadas, e julgar a viabilidade de implementação do sistema proposto. Os requisitos aqui apresentados incluem descrições mínimas, no entanto completas, de todas as funções do sistema de modo a suportar as entradas e saídas do sistema (*inputs* e *outputs*). Os parágrafos estão identificados com as iniciais dos módulos a que se referem os requisitos. Na ausência de identificação do parágrafo, indica que o requisito refere-se a questões genéricas ou que se aplicam a todos os módulos, assim apresentados:

- (A) – Módulo de autoria;
- (O) – Módulo de ontologia;
- (IE) – Módulo de extração da informação;
- (D) – Módulo de descoberta de conhecimento.

## H.1 Interfaces externas

Nesta seção apresentamos uma descrição geral das interfaces externas com suas respectivas entradas e saídas do sistema, a partir das informações levantadas na seção H2, considerando o conteúdo e formato dos itens seguintes.

**Tabela 4 – Interfaces externas do ambiente**

Interfaces		Ambiente de Autoria	Ambiente de Ontologias	Ambiente de Descoberta
1.	<b>Itens</b> Descrição	Ambiente de redação e edição de textos conforme funcionalidades existentes hoje nos principais editores disponíveis no mercado	Ambiente de visualização e edição de ontologias, conforme funcionalidades disponíveis pelo editor selecionado pelo usuário	Ambiente de navegação com página de navegador da Web com funcionalidades específicas daquele selecionado pelo autor
2.	Objetivo	Redigir os textos Visualizar <i>links</i> com ontologias Visualizar marcação ontológica Ligar com ambiente de descoberta Ligar com ambiente de ontologias	Navegar pelas ontologias para aquisição de conhecimento Atribuir ao termo a classe e instância equivalentes na ontologia Pesquisar termos correlatos na ontologia Editar novos itens na ontologia	Navegar por fontes de informação correlatas na Web Atribuir <i>link</i> às fontes selecionadas
3.	Origem de entrada	Autor	Textos digitados no editor	Ontologia e textos digitados no editor
4.	Destino da saída	Ambiente de ontologia e de navegação	Bases de dados de ontologias Ontologia do documento	Ontologia do documento <i>Link</i> respectivo do documento
5.	Unidades de medida	Quantidade de <i>links</i> com ontologia e com ambiente de descoberta	Quantidade de Instâncias reconhecidas no documento	Quantidade de <i>links</i> e conceitos reconhecidos no documento
6.	Tempo de execução	Indeterminado	Indeterminado	Indeterminado
7.	Relação com outras entradas e saídas	Ambientes de ontologia e descoberta	Ambientes de autoria e descoberta	Ambientes de autoria e de ontologia
8.	Formato e organização da tela	Figura 19	Figura 19	Figura 19
9.	Formato e organização da janela	Figura 19	Figura 19	Figura 19
10.	Formato de datas	Padrão corrente	Padrão corrente	Padrão corrente
11.	Formato dos comandos	Padrão corrente	Padrão corrente	Padrão corrente
12.	Mensagens de término (exemplos)	Deseja validar os <i>links</i> existentes com ontologias? Deseja validar os <i>links</i> existentes com fontes de informação? Etc.	Deseja validar o <i>link</i> com a ontologia? Deseja inserir novo item na ontologia? Etc.	Deseja validar <i>link</i> com a URL? Deseja buscar novas fontes de informação? Etc.

## **H.2 Funções do software**

Os requisitos de funcionalidade do ambiente de autoria apresentam as ações fundamentais que o sistema deve desempenhar de forma a aceitar e processar as entradas e saídas do sistema. De acordo com a norma IEEE Std 830-1998, eles devem ser apresentados na regência: “o sistema deve...” e “o autor deve poder...”, para fins de identificação de ator, ação e objeto em questão (entrada ou saída).

A organização dos requisitos ora descritos foi feita com base no modelo de arquitetura, apresentado na Figura 19, da seção B desta especificação de requisitos.

As respectivas funcionalidades dos ambientes estão distribuídas em :

1. Configuração do Ambiente
  - 1.1 Configuração do ambiente de autoria
  - 1.2 Configuração do ambiente de ontologia
  - 1.3 Configuração do ambiente de descoberta
2. Uso do Ambiente
  - 2.1 Seleção das ontologias para uso
  - 2.2 Uso do ambiente de autoria
  - 2.2 Uso do ambiente de extração da informação
  - 2.3 Uso do ambiente de ontologias
  - 2.4 Uso do ambiente de descoberta

### **H.2.1 Configuração do Ambiente**

#### **H.2.1.1 Configuração do ambiente de autoria**

(A) O autor deve poder utilizar um ambiente de editor de texto que ofereça as funcionalidades disponíveis, específicas de cada editor, para formatação do documento.

(A) O autor deve poder ser autenticado para acessar o sistema, informando o nome de usuário e senha na tela de login.

(A) O autor deve poder escolher o software de edição de texto ou de autoria que deseja utilizar.

(A) O sistema deve oferecer a opção de configuração do ambiente de trabalho em: ambiente aberto de autoria – onde não há restrições quanto ao uso das funcionalidades disponibilizadas pelo sistema; e o ambiente fechado de autoria – onde o autor terá restrições no acesso a algumas funcionalidades.

#### **H.2.1.2 Definição do tipo de documento**

(A) O autor deve poder escolher o tipo de formato de documento que será criado, considerando DTDs (Document Type *Description*) que identifiquem se o documento é um artigo, um relatório, uma carta, uma dissertação, uma tese, um livro, etc.

(A) O autor deve poder selecionar os itens que devem constar no documento como título, subtítulo, introdução, apresentação, descrição do problema, objetivos, metodologia, desenvolvimento, conclusão, etc.

(O) O sistema deve reconhecer os itens do documento identificados pelo autor e criar os metadados respectivos aos itens.

(A) O sistema deve permitir que o autor faça a atribuição de tipo de documento e de itens antes e durante a autoria.

### **H.2.1.3 Configuração do ambiente de ontologias**

(O) O sistema deve oferecer a lista de editores e visualizadores de ontologias compatíveis com o sistema.

(O) O autor deve poder selecionar o editor de ontologia a ser utilizado durante a manipulação das ontologias no sistema.

(O) O autor deve poder escolher a forma de visualização de ontologias, em texto (código fonte) ou gráfica, para realizar a atribuição de termos às suas respectivas instâncias na ontologia (ex: em árvore, mapas de cluster, em texto via tópicos, dentre outros).

(O) O sistema deve permitir a alteração do modo de visualização de ontologias durante o processo de autoria.

(O) O sistema deve oferecer ao autor a lista de modelos de metadados disponíveis para representar o documento.

(O) O autor deve poder selecionar o modelo de metadados que visa utilizar como base para a representação do documento.

### **H.2.1.4 Configuração do ambiente de descoberta**

(A) O sistema deve oferecer a lista de navegadores (*browsers*) do ambiente de descoberta que o autor poderá utilizar para navegação semântica.

(D) O autor deve poder escolher a forma de navegação em ambientes de descoberta, se via ontologia ou se via browser.

(A) O autor deve poder selecionar o ambiente de descoberta que deseja utilizar, considerando as fontes de informações relacionadas ao documento, como bases de dados locais ou remotas, interligadas às ontologias existentes.

(D) O sistema deve oferecer ao autor uma listagem de outras fontes de informação existentes, mesmo que não estejam associadas às ontologias (ex: baseadas em sintaxe), para que o autor possa escolher e conseqüentemente associar semanticamente essas outras fontes à ontologia em uso e ao documento em produção.

## H.2.2 Uso do ambiente

### H.2.2.1 Módulo de autoria

(A) O autor deve poder informar os assuntos principais do documento em tela.

(A e O) O autor deve poder selecionar diretamente das ontologias os assuntos principais do documento.

(A e O) O sistema deve poder associar os principais assuntos do documento digitados em tela às respectivas instâncias encontradas nas ontologias em uso.

(A e O) O autor deve poder validar a associação semântica entre os termos digitados em tela e as instâncias das ontologias em uso.

(A e O) O sistema deve poder pesquisar junto a outros BDs externos de ontologias, termos que não foram reconhecidos nas ontologias locais em uso.

(A e O) O sistema deve informar ao usuário a inexistência de conceito equivalente na ontologia, caso não encontre nenhuma relação de equivalência entre o termo digitado e as ontologias disponíveis.

(A e O) O autor deve poder solicitar ajuda ao onto-broker para pesquisa de outras ontologias que possam ser associadas ao documento em produção.

(A e IE) O autor deve poder validar a extração da informação do texto.

(A e IE) O extrator (parser) deve reconhecer os termos digitados pelo autor e buscá-los na ontologia.

(IE, O e A) Após a validação da extração pelo autor, o extrator deve gerar um *link* entre o termo validado e a instância equivalente na ontologia.

(A e IE) O autor deve poder anotar e associar as imagens e demais recursos do documento aos conceitos respectivos na ontologia.

(A) O autor deve poder usufruir de um ambiente de gerenciamento de documentação, geralmente oferecido pelo sistema operacional, com gerenciadores de arquivos e criação de *links* entre os documentos.

(A) O sistema deve oferecer suporte de uso a um ou mais autores simultaneamente, por documento.

(A) O sistema deve oferecer suporte às estruturas lexicais possíveis, por meio de mecanismos como por exemplo tesouros ou dicionários em linha.

(A) O sistema deve oferecer suporte para a criação de gráficos e tabelas.

(A) O sistema deve oferecer mecanismos automáticos para a formatação do documento, no que se refere a layout, titulação, margens, etc.

(A) O sistema deve oferecer opção de restrição de entrada de dados em ambiente fechado de autoria.

(A) O sistema não deve restringir a entrada de dados, em ambiente aberto de autoria.

(A) O sistema deve interagir com o autor, informando ações disponíveis e respectivas restrições de uso das funcionalidades.

(A e D) O ambiente de autoria deve estar ligado ao ambiente de pesquisa e navegação semântica de dados no ambiente de descoberta.

(A e D) O ambiente de autoria deve permitir a navegação em fontes de informação externas a partir dos termos anotados.

(A e O) O ambiente de autoria deve estar ligado ao ambiente de ontologia para prover visualização de interligação semântica entre os conceitos reconhecidos nas ontologias.

(A e O) O sistema deve permitir a criação e edição de formatos de arquivos específicos ao ambiente de produção, como html e xml.

(A e IE) O autor deve poder digitar livremente termos no ambiente de autoria.

(A e IE) O sistema deve realizar o parseamento automático do texto digitado.

(A e O) O sistema deve fazer o reconhecimento da semântica dos termos digitados, a partir da comparação com as ontologias associadas pelo autor.

(A e O) O sistema deve identificar um termo digitado e associá-lo à respectiva classe ou instância semanticamente equivalente, presente na ontologia em uso.

(A e O) O sistema deve associar os termos do documento com as respectivas classes e instâncias reconhecidas das ontologias de domínio ou outras ontologias equivalentes.

(A, IE e O) Os termos digitados e reconhecidos pelas ontologias devem ser marcados e tornar-se *links* do ambiente de autoria para a classe ou instância respectiva na ontologia.

(A e O) O sistema deve oferecer ao autor a visualização da parte da ontologia em que está contido o termo ou conceito digitado.

(A e O) Os termos ligados às ontologias herdam a estrutura da ontologia em uso e automaticamente o sistema deve compor a ontologia do documento em paralelo.

(A) O sistema deve solicitar ao autor a especificação do idioma utilizado no documento.

(A) O sistema deve solicitar ao autor a especificação da linguagem de processamento das informações.

(A) O ambiente deve dar suporte à produção de documentos compostos que utilizam de diversas tecnologias.

(A e IE) O autor deve poder digitar termos aleatórios, frases completas ou também palavras-chave, os quais são parseados e sua estrutura sintática reconhecida.

(IE, O e A) Após a validação da extração pelo autor, o extrator deve gerar um *link* entre o termo validado e a instância equivalente na ontologia.

(A e O) O autor deve validar a associação feita pelo sistema entre o termo digitado e a respectiva classe ou instância na ontologia.

(A e O) O sistema não deve obrigar a validação da associação do termo e ontologia feita pelo autor.

(A) O sistema deve ser de fácil uso e permitir que os usuários expressem conceitos mesmo sem conhecimento aprofundado em representação do conhecimento.

(A) O sistema deve permitir visões do documento em modo espacial, não linear, linear e formatada, conforme funcionalidades do editor de texto em uso.

(A) O sistema deve permitir fácil movimentação entre as visões do documento, conforme funcionalidades do editor de texto em uso.

(A) O sistema deve permitir realizar operações básicas de edição, incluindo a edição do texto, da rede de notas e da estrutura do documento.

(A) O sistema deve permitir a criação de marcações informais e rascunhos; *link* entre o texto do documento a um meta-texto (como anotações, por exemplo).

(A) O sistema deve permitir a configuração e a manipulação de restrições.

(A) O sistema deve permitir a sobreposição de tarefas e incorporação de outras, conforme funcionalidades do editor de texto em uso, como criação de notas e envio de e-mail durante o processo de autoria.

(A) O sistema deve permitir a reutilização de produtos anteriores, incluindo o texto, estrutura do documento e de notas.

(A) O sistema deve permitir a comunicação entre autores colaboradores por uma rede de debate e revisão de texto, assim como entre os engenheiros de ontologia e ontobrokers.

(A) O sistema deve realizar o controle de versão dos documentos gerados.

(A) O sistema deve oferecer opções de mudanças de intenção, incluindo uma grande facilidade de “desfazer” ações.

(A) O sistema deve permitir a junção de múltiplas contribuições de construção e revisão de texto.

(A) O sistema deve permitir o acesso ao histórico de interações entre os autores e o texto.

(A) O sistema deve permitir a alternância de autores participantes do processo.

(A) O sistema deve permitir a realização das 5 etapas de produção do documento: o planejamento, a composição, a validação e a manutenção.

#### **H.2.2.2 Seleção da ontologia a ser usada**

(O) O autor deve poder selecionar as ontologias existentes, as quais deseja utilizar na autoria do documento. Duas ou mais ontologias podem ser utilizadas simultaneamente durante a produção.

(O) O autor deve poder atribuir relação hierárquica entre as ontologias: primária, secundária e terciárias, conforme o grau de abordagem da informação na autoria.

(O) O sistema deve oferecer a opção de configuração de outras ontologias durante a autoria, para aumentar a capacidade de raciocínio ontológico em situações em que haja a expansão da abordagem para outros assuntos que não aqueles definidos no início da autoria.

(O) O sistema deve permitir que sejam associadas ontologias de âmbito local, que tenham sido criadas no ambiente em questão, ou também, outras ontologias de domínio ou globais que se encontram em outros repositórios locais ou remotos.

(O) O sistema deve oferecer ao autor a lista de ontologias existentes para acesso imediato, assim como outras fontes de ontologias correlatas que podem ser solicitadas e baixadas para o ambiente de produção.

(O) O sistema deve oferecer a opção de configuração do ambiente ontológico antes e durante a autoria, pois novos assuntos podem ser incluídos durante o processo de produção.

(O) A autor, caso não encontre as ontologias, deve poder pesquisar novas ontologias na Web ou solicitar ao ontobroker.

(O) O autor deve poder selecionar o formato e a linguagem da ontologia a ser utilizada.

(O) O sistema deve converter a ontologia do formato definido pelo autor para o formato padrão definido pelo W3C.

### H.2.2.3 Módulo de Ontologias

- (O) O sistema deve permitir a edição de ontologias em OWL e baseadas em esquema RDF.
- (O) O sistema deve oferecer ao autor a visualização da parte da ontologia em que está contido o termo ou conceito digitado.
- (O) Os termos ligados às ontologias herdam a estrutura da ontologia em uso e automaticamente o sistema deve compor a ontologia do documento em paralelo.
- (O) O sistema deve permitir a visualização de toda a ontologia, por partes específicas.
- (O) O autor deve poder incluir um novo termo na ontologia existente, relacionando-o a classes e instâncias das ontologias em uso definidas no ambiente, para a edição de novos conceitos na ontologia.
- (O) O sistema deve aceitar as novas edições de ontologia feitas pelo autor.
- (O) O sistema deve submeter ao engenheiro de ontologias as novas edições de ontologia feitas pelo autor, para aprovação e decisão de inclusão na ontologia do ambiente.
- (O) O sistema deve oferecer opção de visualização da ontologia do documento.
- (A e O) O autor deve poder solicitar ajuda ao ontobroker para pesquisa de outras ontologias que possam ser associadas ao documento em produção.
- (A e O) O autor deve poder começar a utilizar o ambiente de edição a partir das ontologias. Ao navegar, pelas mesmas, pode clicar e arrastar o termo da ontologia para dentro do documento, constituindo assim numa forma inversa de fazer a anotação semântica.
- (A e O) O autor deve poder redigir o texto considerando os termos incluídos da ontologia, os quais já estão anotados e referenciados ao ambiente de descoberta.
- (A e O) O autor deve poder validar a associação do termo na ontologia, podendo incluir novo e editar relação entre o termo e respectiva instância.
- (A e O) O ambiente de autoria deve estar ligado ao ambiente de ontologia para prover visualização de interligação semântica entre os conceitos reconhecidos nas ontologias.
- (A e O) O sistema deve fazer o reconhecimento da semântica dos termos digitados, a partir da comparação com as ontologias associadas pelo autor.
- (IE e O) O sistema deve buscar na ontologia os termos reconhecidos no extrator/parser.

(IE e O e D) O sistema deve tornar o *link* reconhecido e anotado pela ontologia com as bases de informação existentes do ambiente de descoberta, para a navegação semântica.

(A e O) O sistema deve oferecer ao autor a visualização da parte da ontologia em que está contido o termo ou conceito digitado.

(O) Caso os termos digitados não tenham sido reconhecidos diretamente pela ontologia, o sistema deve apresentar na ontologia o conceito que mais se aproxima ao termo digitado (semântica ou sintaticamente).

(O) Se houver o conceito equivalente na ontologia (via pesquisa do autor), o sistema deve permitir que o autor valide a associação a outros conceitos.

(O) Se o conceito ou equivalente não estiver na ontologia, o autor deve poder criar um novo conceito na ontologia, inserir o termo, editar as relações com os termos existentes ou apenas manter o termo novo isolado. Essa ação de inserção será enviada ao banco de dados de ontologias para validação pelo engenheiro de ontologias.

(O) O onto-broker deverá solicitar a busca, em outras ontologias, do novo conceito sugerido pelo autor.

(O) Caso não haja ontologia disponível, o ontobroker deverá solicitar a criação da nova ontologia específica, equivalente ao conceito novo sugerido pelo autor.

(O) O sistema deve anotar e marcar como *hiperlink* para a ontologia equivalente, os termos digitados pelo autor e reconhecidos na ontologia.

(O) O sistema deve oferecer ao autor a lista de modelos de metadados disponíveis para representar o documento.

(O) O autor deve poder selecionar o modelo de metadados que visa utilizar como base para a representação do documento.

(O) O autor deve poder associar às categorias de metadados aos termos principais equivalentes do documento.

(O) O autor deve poder validar a ontologia do documento sugerida pelo sistema.

(O) O autor deve poder editar a ontologia do documento conforme suas necessidades

(O) O autor deve poder submeter a ontologia do documento para o engenheiro de ontologias para fins de revisão e validação.

(O) O engenheiro de ontologias deve poder revisar e validar a ontologia do documento criada pelo sistema e submetida pelo autor.

(O e D) O sistema deve tornar a ontologia do documento em uma estrutura de *links* às fontes de informação disponíveis no ambiente de descoberta.

(O e D) O autor deve poder associar fontes de informação correlatas à ontologia do documento criada.

(O e D) O autor deve poder validar a associação de fontes de informação sugeridas pelo sistema no ambiente de descoberta.

(O e D) O autor deve poder indicar novas fontes de informação disponíveis, e editar endereços (URLs) novos para associação às fontes.

(A e O) O autor deve poder validar o assunto associado automaticamente nas classes ou instâncias das ontologias.

(A e O) O autor deve poder validar a associação automática realizada pelo sistema entre os assuntos principais do texto e as ontologias apresentadas.

(O) O autor deve poder incluir ou editar novos termos e conceitos na ontologia, caso não existam.

(O) O autor deve poder validar novas inserções na ontologia junto ao engenheiro de ontologias.

(O) O autor deve poder finalizar o documento, submetendo-o para validação ontológica, junto ao engenheiro de ontologias.

(A e O) O sistema deve reconhecer o termo digitado no documento a partir das ontologias.

(O) O sistema deve submeter os termos não encontrados ao autor, para decisão de inclusão na ontologia ou de submissão de novas ontologias ao ontobroker.

(O) O onto-broker deve pesquisar junto a bases locais e externas, à busca do termo novo não reconhecido.

(IE e O) O sistema deve anotar e ligar o termo reconhecido à respectiva ontologia encontrada.

(IE e O) O sistema deve extrair o termo e pesquisá-lo na ontologia.

(IE e O) O sistema deve extrair o termo digitado e pesquisar na ontologia a sua posição.

(O) O sistema deve oferecer lista de padrões de metadados existentes.

(O) O autor deve poder selecionar o formato de metadados que deseja utilizar.

(O) O sistema deve reconhecer o termo digitado com os formatos de metadados associados.

(O) O sistema deve oferecer a categorização dos termos reconhecidos com as possíveis categorias de metadados, de acordo com o formato definido pelo autor.

(O) O sistema deve perguntar ao autor se associação automática está adequada.

(O) O autor deve poder alterar a associação de metadados às categorias, sugerindo outros termos ou alterar categorias.

(O) O autor deve poder validar a estrutura de metadados sugerida e/ou alterada.

(O) O autor deve poder acessar e navegar livremente pela estrutura ontológica para adquirir conhecimento sobre um conceito.

(O) O sistema deve oferecer mecanismo de validação ontológica do documento, comparada às estruturas ontológicas existentes, definidas pelo autor para uso corrente durante a autoria.

(O) O autor deve poder localizar o conceito de uma área ou domínio e certificar-se de que é o adequado ao uso.

(O) O autor deve poder optar por usar um termo variante, desde que esteja presente na ontologia relacionado ao termo principal, em ambiente fechado de autoria.

(A e O) O sistema deve sugerir os termos preferidos para serem usados em ambientes fechados de autoria.

(A e O) O sistema deve oferecer a opção de obrigatoriedade ao uso de termos preferidos em ambientes fechados de autoria.

(A e O) O sistema deve oferecer a liberdade de uso de termos variantes em ambientes abertos de autoria.

(O e D) O sistema deve oferecer a opção de pesquisa de fontes de informação em ambiente de descoberta a partir das instâncias equivalentes da ontologia.

(IE, O e A) Após a validação da extração pelo autor, o extrator deve gerar um *link* entre o termo validado e a instância equivalente na ontologia.

(IE e O) O sistema deve anotar o termo digitado e associá-lo a sua respectiva instância na ontologia.

(IE e O) O sistema deve anotar e ligar o termo reconhecido é anotado à ontologia encontrada.

(O) O sistema deve compor a ontologia específica do documento, com base no reconhecimento de entidades presentes na ontologia em uso, adotando a mesma relação hierárquica dos termos do documento que a dos conceitos presentes nas ontologias em uso.

- (O) O sistema deve sugerir o uso de outras ontologias, a partir do momento que reconhece o uso de outras entidades conceituais que não estão presentes na ontologia em uso.
- (O) O sistema deve oferecer uma lista de padrões de metadados existentes para fins de representação da estrutura de conteúdo do documento, como por exemplo DC (Dublin Core), MARC, CDU (Classificação Decimal Universal), etc.
- (O) O sistema deve prover uma estrutura de conceitos baseada em tesauro para a identificação de relações semânticas entre os termos.
- (O) O sistema deve prover a visualização de estruturas ontológicas associadas.
- (O) O sistema deve prover a comparação entre a ontologia do documento recém-criada com ontologias existentes, por meio de mecanismos de visualização gráfica ou textual, permitindo a verificação de consistência entre elas.
- (O) O sistema deve oferecer aos autores a opção de criação da ontologia do documento, de modo que essa ontologia possa ser associada a outras ontologias globais ou de domínio que foram especificadas pelo autor.
- (O) O sistema deve oferecer um mecanismo automático de checagem de inconsistência e ambigüidade na ontologia em uso, por meio de um mecanismo de alerta ao autor de que há algum tipo de inconsistência detectada.
- (O) A atividade de engenharia ontológica deve ser realizada durante a autoria do documento, automaticamente pelo sistema, apoiada por validação via um engenheiro de ontologias.
- (O) O autor deve poder popular a ontologia em uso durante o processo de autoria para completar a abrangência da estrutura semântica com os conceitos novos apresentados e que não estejam na ontologia original.
- (O) O autor deve poder navegar pelas ontologias, a partir de mecanismos de visualização em formatos diversos durante a pesquisa semântica concomitante à autoria do documento.
- (O) O autor deve poder interrelacionar as ontologias disponíveis para aumentar a capacidade de pesquisa e de representação semântica do documento.
- (O) O autor deve poder checar via ontologia a existência de ambigüidade, inconsistência ou variação terminológica dos termos utilizados no documento, por meio das ontologias ou de fontes de informação referenciadas.
- (O) O sistema deve oferecer uma representação semântica detalhada do documento, com a maior quantidade possível de conceitos reconhecidos pelas ontologias ou indicados pelo autor durante a autoria.

- (O) O sistema deve oferecer um guia de visualização da ontologia, para facilitar a sua compreensão e identificação de instâncias relacionadas aos conceitos reconhecidos.
- (O) O sistema deve oferecer uma meta-ontologia que situe os principais assuntos relacionados ao documento.
- (O) O sistema deve fazer checagem automática da existência dos termos e conceitos na ontologia por meio de um *crawler*, que realiza a busca na ontologia por meio de agentes com raciocínio ontológico.
- (O) As ontologias devem fazer inferência automática de conteúdo, e interagir com o autor para validar ou complementar a inferência realizada.
- (O) O sistema deve oferecer a opção de movimentação da ontologia para a visualização de todas as relações existentes com o conceito em uso.
- (O) A movimentação da ontologia deve ser feita de forma rápida e de fácil visualização.
- (O) As opções do ambiente de ontologia devem estar organizadas em módulos integrados que facilitem a execução de ações específicas como visualização, atribuição de conceitos relacionados, validação de uma associação semântica, etc.
- (O) A interface de visualização de ontologias deve permitir a clara identificação de conexões e relações semânticas entre os conceitos.
- (O) O ambiente deve permitir a visualização de uma ou mais ontologias simultaneamente.
- (O) As ontologias utilizadas simultaneamente devem poder ser unidas automaticamente para facilitar a identificação de similaridade ou diferença entre as mesmas.
- (O) As ontologias devem ser alinhadas entre si para permitir a precisão semântica oriunda da união e integração entre as mesmas.
- (O) O sistema deve ter um mecanismo de controle de versão das ontologias utilizadas para permitir o rastreamento das modificações realizadas.
- (O) A disponibilização de ontologias no ambiente de produção deve oferecer suporte a contextos específicos, com ontologias de domínio e ontologias globais relacionadas.
- (O) As ontologias em uso devem prever mecanismo de raciocínio ontológico para o reconhecimento de ambigüidades e resolução de problemas de inconsistência.
- (O) O autor deve poder editar/popular a ontologia em uso, de acordo com suas necessidades específicas.

(O) O engenheiro ontológico deve analisar e validar as edições e populações ontológicas realizadas pelo autor.

(D e O) O sistema deve importar para a ontologia do documento a estrutura de *links* definidos e que foram associados à ontologia.

(O) O autor deve poder atribuir termos da ontologia do documento às categorias de metadados, criando um esquema independente de metadados do documento.

(IE, O e A) Após a validação da extração pelo autor, o Extrator deve gerar um *link* entre o termo validado e a instância equivalente na ontologia.

#### **H.2.2.4 Módulo de Extração da Informação**

(A e IE) O sistema deve reconhecer os termos digitados pelo autor e buscá-los na ontologia.

(A e IE) O sistema deve permitir ao usuário digitar livremente termos no ambiente de autoria e realizar o parseamento dos termos automaticamente.

(IE) O sistema deve fazer o reconhecimento de entidades com nomes próprios (NER).

(A e IE) O autor deve poder validar a extração da informação do texto.

(IE) O sistema deve informar ao autor o parseamento realizado.

(IE) O autor deve poder validar o parseamento realizado pelo sistema.

(IE, O e A) Após a validação da extração pelo autor, o extrator deve gerar um *link* entre o termo validade e a instância equivalente na ontologia.

(IE e O) O sistema deve extrair o termo e pesquisá-lo na ontologia.

(IE e O) O sistema deve anotar e ligar o termo reconhecido à respectiva ontologia encontrada.

(IE e O) O sistema deve extrair o termo digitado e pesquisar na ontologia a sua posição.

(IE, O e A) Após a validação da extração pelo autor, o extrator deve gerar um *link* entre o termo validado e a instância equivalente na ontologia.

(IE) A anotação semântica da informação deve ser o mais automático possível.

(IE e O) O sistema deve anotar e ligar o termo reconhecido e anotado à ontologia encontrada.

(IE) O sistema deve solicitar ao usuário a escolha de anotação semântica automática ou sob demanda.

(IE e O) O sistema deve buscar na ontologia os termos reconhecidos no extrator/*parser*.

(IE e O) O sistema deve apresentar ao autor os termos que não foram reconhecidos na ontologia, para sugerir a edição da ontologia.

(IE e O e D) O sistema deve tornar o *link* reconhecido e anotado pela ontologia com as bases de informação existentes do ambiente de descoberta, para a navegação semântica.

(IE) O autor deve poder homologar e estruturar os *links* com as referências semânticas do ambiente de descoberta.

(A e IE) O autor deve poder anotar e associar as imagens e demais recursos do documento aos conceitos respectivos na ontologia.

(IE) O sistema deve fazer o parseamento do texto por meio de análise léxica automática.

(IE) O sistema deve identificar as variações terminológicas no texto, por meio do uso de tesouros durante o parseamento.

(IE) O sistema deve apresentar ao usuário a lista de termos variantes encontrados no texto a partir do tesouro.

(IE) O sistema deve oferecer ao autor a opção de normalização da estrutura sintática manualmente, identificando sujeito, verbo, predicado, objetos e complementos.

(IE) O sistema deve apresentar os casos de ambigüidade ao autor para validação.

(IE) O sistema deve gerar a informação sintática final – estrutura da frase.

(IE e O e D) O sistema deve anotar e ligar os termos reconhecidos à ontologia e ao ambiente de descoberta.

(IE e O) O sistema deve marcar e anotar o termo presente na ontologia.

(IE) O sistema deve marcar e anotar o termo aprovado pelo autor.

(IE e O) O sistema deve marcar os termos no documento que foram reconhecidos pela ontologia.

(IE e O) Os termos reconhecidos pelas ontologias são extraídos e compõem uma ontologia específica do documento.

(IE e O) O autor deve poder anotar os conceitos em uso no documento, relacionando-os a classes e instâncias da ontologia ou a novas instâncias por ele criadas.

(IE) O sistema deve prover anotação semântica automática ou semi-automática do texto.

### **H.2.2.5 Módulo de descoberta de conhecimento**

(A e D) O ambiente de autoria deve estar ligado a ambiente de pesquisa e navegação semântica de dados no ambiente de descoberta.

(A e D) O ambiente de autoria deve permitir a navegação em fontes de informação externas a partir dos termos anotados.

(O e D) O sistema deve tornar a ontologia do documento em uma estrutura de *links* às fontes de informação disponíveis no ambiente de descoberta.

(D e O) O sistema deve importar para a ontologia do documento a estrutura de *links* definidos e que foram associados à ontologia.

(O e D) O autor deve poder associar fontes de informação correlatas e existentes à ontologia do documento criada.

(O e D) O autor deve poder sugerir novas fontes de informação correlatas e novas à ontologia do documento criada.

(O e D) O sistema deve herdar a lista de fontes de informação associadas às ontologias em uso para a ontologia do documento.

(O e D) O autor deve poder validar a associação de fontes de informação sugeridas pelo sistema no ambiente de descoberta.

(O e D) O autor deve poder indicar novas fontes de informação disponíveis, e editar endereços (URLs) novos para associação às fontes.

(O e D) O autor deve poder submeter o documento para revisão de pares e editora, por meio de sistema de transferência de arquivos (como e-mail, ftp, upload, etc.).

(D) O autor deve poder definir, durante a navegação semântica, quais fontes e/ou partes do texto encontrados no ambiente de descoberta são âncoras semânticas para o termo respectivo no documento original.

(D e O) O sistema deve importar para a ontologia do documento a estrutura de *links* e referências a fontes de informação definidos e que foram associados à ontologia.

(D) O sistema deve associar termos não reconhecidos de forma sintática a fontes de informação presentes no ambiente de descoberta.

(O e D) O sistema deve oferecer a estrutura de navegação via ontologia, ligada ao ambiente descoberta.

(D) O sistema deve oferecer a estrutura de navegação semântica para que o autor possa criar um *link* entre os termos e as páginas-alvo.

(D) O autor deve poder editar e validar os *links* entre o termo e o ambiente de descoberta.

(D) No caso de termo reconhecido pela ontologia, cria-se a referência automaticamente, e faz-se o *link* direto com as bases já representadas semanticamente.

(D) O sistema deve prover um ambiente de navegação semântica via ontologias, interligado a fontes de informação locais e remotas.

(D) O autor deve poder acessar fontes de informação disponíveis local ou remotamente por meio de ontologias.

(D) O autor deve poder acessar fontes de informação disponíveis localmente ou remotamente por meio dos *links* nos termos do documento.

(D) O sistema deve permitir anotação e referência de fontes de informação externas.

(D) O sistema deve apresentar a relação decorrente do reconhecimento do termo na ontologia, para que o autor possa validar a relação.

(D) O autor deve poder interferir na representação da informação em ambientes de autoria abertos.

(D) O autor deve poder solicitar ao sistema que a representação da informação seja feita de forma automática.

(D) O sistema deve solicitar ao autor a validação da representação automática da informação realizada.

(D) O sistema deve oferecer ao autor a navegação semântica por meio do visualizador de ontologia, por meio de *links* entre os termos e bases de dados locais e a Web.

(D e O) A ontologia do documento deve ser utilizada para navegar em ambientes de descoberta relacionados aos conceitos nela presentes.

(D e O) A ontologia do documento deve herdar os *links* pré-existentes entre a ontologia em uso e o ambiente de descoberta.

(D e O) O autor deve poder validar os *links* da ontologia do documento herdado da ontologia em uso e relacionadas ao ambiente de descoberta.

## **I. Requisitos de Desempenho**

Nesta seção, apresentamos alguns requisitos básicos de desempenho do ambiente, com base na necessidade das funções dos módulos, e tendo como referência as orientações levantadas na análise dos requisitos realizada no Capítulo 3. Considerando-se que a proposta do ambiente de autoria apresentada encontra-se em estado de especificação de alto nível, ainda não é possível mensurar o volume e tempo de transações para garantir o desempenho desejado. Para tanto, sugere-se, em estudos futuros que sejam feitas análises para mensuração da desempenho a partir de um protótipo do sistema, para que uma projeção mais apurada possa ser realizada, e que se possa constatar precisamente e em detalhes os requisitos de desempenho necessários.

No entanto, apresentamos abaixo alguns requisitos que podem ser levados em consideração para se definir a quantidade suficiente de memória e robustez do ambiente que possa viabilizar a solução proposta.

- O sistema deve permitir o uso de um ou mais usuários simultaneamente.
- O sistema deve solicitar autenticação dos usuários e armazenar log de acessos.
- O sistema deve dar suporte à autoria colaborativa de documentos.
- O sistema deve oferecer níveis de acesso aos usuários, de acordo com as funcionalidades disponibilizadas, em níveis de autoria, edição, revisão e consulta, por exemplo.
- O sistema deve permitir o uso de uma ou mais ontologias, simultaneamente, durante a autoria do documento.
- O sistema deve permitir a visualização das ontologias com facilidade e rapidez, e permitir a fácil alternância entre os modos de visualização.
- O sistema deve permitir o processamento das informações no ambiente de autoria com rapidez e precisão.
- O sistema deve possuir memória de processamento suficiente para realizar a extração automática da informação.
- O sistema deve apresentar característica de robustez suficiente para processar o volume de informações, o uso de formatos de arquivos diferentes, a conversão de formatos e ontologias, a comunicação entre os atores, assim como a visualização da informação em ambiente de descoberta.

## **J. Conformidade com padrões do W3C**

Nesta seção são especificados os requisitos oriundos de padrões e regulamentações existentes e que dão base para a construção do ambiente ora proposto. Considerando-se a referência principal do trabalho que é a Web Semântica, são apresentados nesta seção requisitos constantes das recomendações, trabalhos em andamento (*working drafts*) e notas dos grupos de trabalho que se referem a ambientes de autoria, linguagem de autoria e linguagens de ontologias (W3C, 2006).

São apresentadas adaptações das recomendações do W3C que beneficiam a criação de um ambiente de autoria baseado em ontologias. Os requisitos

selecionados podem ser adotados na íntegra de modo a facilitar a adequação do ambiente às melhores práticas e recomendações dos grupos de trabalho do W3C, listados a seguir.

- Requisitos para autoria de documentos em XML
- Requisitos para uso de ontologias em OWL
- Requisitos para autoria de documentos compostos
- Requisitos para acessibilidade do ambiente de autoria
- Requisitos para autoria independente de dispositivos

Para adequação das linguagens HTML e XHTML, o W3C sugere total conformidade com as recomendações geradas (W3C, 1999, 2000, 2005), as quais devem ser seguidas de acordo com os requisitos e regras de sintaxe definidos, e o mesmo ocorre com o RDF (W3C, 2004).

### ***J.1 Requisitos para documentos em XML***

Nesta seção apresentamos os requisitos específicos para a autoria de documentos em XML, adaptados das seguintes recomendações do W3C: XML Schema Requirements (W3C, 1999) e XML Processing Model Requirements and Use Cases (2006)

A especificação de parâmetros deve conter nomes padronizados para os componentes correspondentes, como por exemplo:

- XML Base
- XInclude
- XSLT 1.0/2.0
- XSL FO
- XML Schema
- XQuery
- RelaxNG
- Dentre outros (para conceitos, vide W3C, 2006).

Os parâmetros do XML devem permitir que as aplicações definam e compartilhem novos passos que usem componentes novos ou já existentes (citados no requisito acima).

O sistema deve conter um inventário mínimo de componentes definidos pela especificação que são necessários para dar suporte e facilitar a interoperabilidade de parâmetros XML.

O sistema deve prover mecanismo para a composição de parâmetros XML para reutilização em outros propósitos, fornecidos por meio de um documento de especificação de parâmetros XML.

Os parâmetros XML devem permitir a iteração entre um conjunto específico de passos e a coleção de documentos ou os elementos de um documento.

Os parâmetros XML devem oferecer mecanismos para o processamento condicional de documentos ou elementos específicos dentro do documento, que sejam baseados em avaliação de expressão, oferecidos pelo grupo de trabalho de requisitos de modelagem de XML do W3C.

Os parâmetros XML devem oferecer mecanismos para identificar erros e comportamentos indesejados no ambiente de autoria.

Os parâmetros XML devem dar suporte ao modelo de dados XPath 2.0.

Um parâmetro de XML deve permitir implementações mais sofisticadas nos documentos e permitir a realização de operações paralelas sem desgaste da capacidade de processamento do ambiente, assim como outras otimizações.

Um parâmetro de XML deve permitir a existência de parâmetros-guia para certas instâncias como uma opção de otimização eficiente da parametrização de um documento XML.

Os documentos em XML devem começar informando no cabeçalho do documento a versão específica do XML que está sendo usada.

(A) Os documentos em XML devem ser legíveis aos humanos e devem ser claros e fáceis de serem interpretados.

(A) O design de um documento deve ser formal e conciso e fácil de ser criado.

(A) O sistema deve permitir a criação de documentos em XML bem formados, ou seja, para cada elemento criado, deve haver uma tag de abertura e uma de fechamento.

(A) O sistema deve atribuir a cada entidade (termo ou seqüência de caracteres) parseada uma marcação ou um caractere específico.

(A) Os comentários adicionados no documento devem aparecer fora das tags e não são considerados caracteres do documento.

(A) O editor de XML deve permitir a recuperação dos comentários do documento.

(A) Os documentos em XML devem começar com uma declaração que especifique a versão do XML em uso (declaração do tipo de documento – DTD), e essa declaração deve aparecer antes do primeiro elemento do documento.

(A) A declaração do tipo de documento deve ser especificada no início do documento XML para definir uma gramática para uma classe de documentos (DTD – Document Type Definition) ou seja, estruturas lógicas ou unidades de armazenagem de informações ou metadados do documento.

(A) Documentos *standalone*, ou seja, que não apresentam marcação externa a outros documentos, devem ser declarados.

(A) Os espaços em branco utilizados no documento devem ser declarados para que sejam reconhecidos pelo processador de XML.

(A) O processador de XML deve reconhecer cada fim de linha como entidade passível de parseamento.

(A) O idioma utilizado para escrever o conteúdo do documento deve ser declarada.

(A) O nome de um elemento constante em uma tag de abertura deve ser o mesmo nome utilizado em uma tag de fechamento equivalente.

(A) A estrutura de um tipo de elemento do documento em XML deve ser declarada.

(A) A especificação de atributos deve aparecer dentro das tags e deve ser declarada por meio de listas de atributos pertencentes a cada tipo de elemento. Servem para restringir os tipos de atributos e também fornecer valor aos atributos.

(A) Cada entidade de um documento em XML (unidades de armazenamento) ou referência deve ser identificada por um nome específico.

(A) O processador de xml deve poder ler cada unidade de informação do texto e reportar a consistência do documento (das entidades e referências criadas) e as violações de validade das tags.

(A) O processador de xml deve poder identificar a codificação de caractere utilizada (*character encoding*).

## **J.2 Requisitos para ontologias em OWL**

As ontologias são utilizadas para aumentar a capacidade de representação do conhecimento de aplicações baseadas em Web e permitirão novas possibilidades de uso da Web Semântica (W3C, 2004). Nesta seção estão descritos requisitos específicos de conformidade com a linguagem OWL, levantados a partir da recomendação *OWL – Web Ontology Language Use Cases*

*and Requirements*<sup>4</sup> do W3C. Os requisitos foram escolhidos por esse grupo de trabalho a partir de aspectos de casos de uso considerados mais importantes e que podem ser utilizados como base para o uso dessa linguagem em ambientes de autoria.

Os casos de uso e as metas de design definidas por essa recomendação do W3C motivaram um número de requisitos essenciais para a linguagem de ontologia da Web (OWL). Cada requisito apresentado segue uma respectiva motivação que deu origem ao padrão estabelecido. As motivações estão disponíveis na recomendação e abrangem as seguintes características: uso de ontologias compartilhadas, interoperabilidade entre ontologias, evolução de ontologias, detecção de inconsistência, facilidade de uso, compatibilidade com outros padrões, internacionalização, expressividade e escalabilidade.

Logo, entende-se que para que seja processado pela Web Semântica, um ambiente de autoria deve considerar os seguintes requisitos no tocante à linguagem das ontologias que serão utilizadas pelos autores.

(O) As ontologias devem ter um único identificador próprio, para garantir a identificação unívoca de cada ontologia em uso (URI).

(O) As ontologias devem referenciar os conceitos por meio de identificadores (URIs) de forma não ambígua. Dois conceitos em diferentes ontologias devem ter identificadores distintos e absolutos.

(O) As ontologias devem poder explicitamente expandir outras ontologias de modo a reutilizar conceitos durante a inclusão de novas classes ou propriedades. Se a ontologia A expande a ontologia B, e a ontologia B expande a ontologia C, então a ontologia A implicitamente expande a ontologia C.

(O) Os recursos devem estar atribuídos ou relacionados (*committed*) a ontologias específicas, indicando precisamente o conjunto de definições e considerações feitas.

(O) O sistema deve prover metadados para cada ontologia utilizada, como autor, data de publicação da ontologia, etc., com base em padrões como o Dublin Core.

(O) A linguagem da ontologia utilizada deve oferecer meios de comparação e relacionamento de diferentes versões de uma mesma ontologia.

(O) A linguagem da ontologia utilizada deve permitir expressar definições complexas de classes, como interseção, união e complementação de classes, por exemplo.

(O) A linguagem da ontologia utilizada deve expressar as definições das propriedades, como por exemplo, sub-propriedades, propriedades inversas, transitividade, limitações do domínio ou range.

---

<sup>4</sup> Casos de uso e requisitos para OWL – Linguagem de Ontologia da Web

(O) A linguagem da ontologia deve prover um conjunto de tipos de dados (*data types*), com base no esquema XML, para garantir a compatibilidade com outros padrões.

(O) A linguagem deve incluir características para estabelecer que duas classes ou propriedades sejam equivalentes, para permitir a interoperabilidade entre ontologias.

(O) A linguagem deve incluir características que permitam estabelecer que dois pares de identificadores representem o mesmo indivíduo (instância), para permitir a interoperabilidade entre ontologias.

(O) A linguagem deve oferecer uma forma de permitir que afirmações adicionais sejam incluídas na ontologia, como por exemplo data, nível de confidencialidade, recurso, etc.

(O) A linguagem deve dar suporte ao tratamento de classes e instâncias, uma vez que um mesmo conceito pode ser visto como uma classe ou indivíduo da ontologia, dependendo da perspectiva do usuário, permitindo a interoperabilidade entre ontologias.

(O) A linguagem deve dar suporte à especificação de restrições de cardinalidade nas propriedades da ontologia, delimitando o número máximo e mínimo de indivíduos a que um outro indivíduo possa se relacionar por meio de uma propriedade específica.

(O) A linguagem deve apresentar uma sintaxe em XML, para permitir a compatibilidade da ontologia com outros padrões.

(O) A linguagem deve suportar a especificação de múltiplas alternativas de rótulos apresentados aos usuários para os recursos especificados na ontologia, por exemplo, visualizar a ontologia em diferentes tipos de linguagem natural.

(O) A linguagem deve suportar o uso de conjuntos de caracteres multilíngües, para permitir a internacionalização e compatibilidade com outros padrões.

(O) A linguagem deve suportar strings de unicode (como funções de teclado, por exemplo).

### ***J.3 Requisitos para Documentos compostos***

A seguir encontram-se os requisitos gerais a serem aplicados em documentos compostos tanto por referência (CDR) quanto por inclusão (CDI). Os requisitos foram adaptados da seguinte recomendação do W3C – Requisitos e Casos de Uso para Documentos Compostos (W3C, 2005).

(A) O formato de um documento composto deve abranger especificações existentes e correntemente em uso que favoreçam as especificações do W3C sempre que possível, e deve limitar a definição de novas marcações, ao menos que

absolutamente requisitadas para fins de integração. Permitir a integração de formatos existentes, e não criar novas formas de marcação.

(A) O formato deve especificar um conjunto base de formatos, correspondendo a perfis e versões específicos.

(A) Os perfis de formato devem definir claramente um critério de conformidade do documento.

(A) Os perfis de formato devem definir claramente o critério de conformidade do agente do usuário (*user agents*<sup>5</sup>).

(A) Cada perfil deve explicar como o agente do usuário (*user agent*) poderá identificá-lo.

(A) Cada perfil deve dar suporte à identificação da marcação e versão em documentos compostos.

(A) Cada perfil deve dar suporte à divulgação das versões específicas de formatos e respectivas capacidades no cabeçalho do documento, como por exemplo, versão, documento tipo mime, etc.

(A) O CDF deve especificar técnicas genéricas de integração, as quais serão aplicáveis em outros grupos de formatos, versões e perfis exigidos pela especificação de um CDF.

(A) O CDF deve dar suporte a mecanismos que estão além da capacidade de especificação de *namespaces*.

(A) O CDF deve dar suporte a linguagens de *script*, as quais devem poder acessar e modificar os elementos tanto de documentos raiz ou documentos-filho assim como as linguagens principais e demais linguagens integradas.

(A) O CDF deve especificar o número máximo de combinações entre linguagens, se por exemplo um documento SVG pode estar referenciado a um outro XHTML.

(A) Os perfis CDF devem prover um método de tratamento de eventos utilizando marcação declarativa para os formatos em uso. O tratamento de eventos deve conter instruções de como lidar com um certo evento, por meio de linguagem de *script* ou uma linguagem declarativa (marcação), tanto em documento-pai ou documento-filho.

(A) O CDF deve definir sua integração com a arquitetura Web. Deve incluir o tráfego de documentos via http e deve buscar ser independente de meios de transferência de arquivos na Web (como RTP, por exemplo).

---

<sup>5</sup> **User Agent** - agente de usuário é uma ferramenta como um browser que apresenta conteúdo.

(A) O CDF deve permitir a compactação de dados, para facilitar o processamento pelos *user agents* e outros navegadores.

(A) O CDF deve permitir que os geradores de conteúdo realizem a descrição ou autoria de conteúdo em mídia enriquecida.

(A) O CDF deve dar suporte à sincronia temporal de conteúdo dinâmico oriundo de múltiplas referências, possivelmente com múltiplas referências a uma mesma fonte de informação, como animação, áudio, vídeo, etc.

(A) O perfil de CDF deve especificar a transversalidade com outros documentos compostos ou linguagens compostas (ex: XHTML referenciando um documento SVG).

(A) O perfil CDF deve especificar *trigger* de animações em todos os *namespaces*.

(A) Os documentos CDF devem atender aos requisitos de acessibilidade.

(A) Os documentos CDF devem dar suporte à atualização dinâmica de marcações, referências, etc.

(A) O CDF deve definir formas com que *soft-keys* e *accesskeys* (atalhos de teclado e outros dispositivos) são utilizados pelos diferentes componentes de linguagens e documentos em uso.

(A) Os agentes de usuários CDF devem prover uma fonte *default* para uso por todos os componentes. Não deve obrigar o uso de uma fonte ou tecnologia particular, mas deve oferecer uma fonte *default*.

(A) O CDF deve permitir adaptações no servidor para viabilizar o processamento de documentos em ambientes com baixa capacidade de memória.

(A) O CDF deve dar suporte a redes de banda limitada ou tecnologias de capacidade limitada, como capacidade de memória, visualização, processamento, etc.

#### ***J.4 Requisitos de Acessibilidade***

Nesta seção, apresentamos os requisitos de acessibilidade do ambiente, com base na recomendação do WAI-TAG do W3C: diretrizes de acessibilidade para ferramentas de autoria para a Web Semântica (W3C, 2005). As recomendações referem-se a usuários que não podem ver, ouvir, mover-se ou ser capaz de processar facilmente alguns tipos de informações, ou não pode processá-las; usuários que tenham dificuldade de leitura ou compreensão de texto; usuários que tenham dificuldade em utilizar teclado ou mouse; que tenham monitores (telas) pequenos ou com visualização textual apenas.

(A) O sistema deve apresentar uma interface acessível a usuários portadores de necessidades especiais.

(A) A interface do sistema deve ser perceptível ao usuário, provendo alternativas de visualização de textos para conteúdos não-textuais (ex: tecnologia assistida ou braile).

(A) O sistema deve permitir que as interfaces sejam configuráveis para o tipo de tecnologia utilizada por usuários com necessidades especiais.

(A) O autor deve poder configurar a forma de apresentação das visões de edição sem alterar o conteúdo que está sendo editado.

(A) O sistema deve permitir que a informação, a funcionalidade e a estrutura possam ser separadas da apresentação do conteúdo.

(A) A interface do sistema deve ser operável, ou seja, o sistema deve permitir que todas as funcionalidades disponíveis possam ser acessadas via teclado ou outra interface de teclado (teclado virtual, por exemplo).

(A) O autor deve poder, via teclado, executar qualquer tarefa de autoria (navegação, seleção ou edição de conteúdo).

(A) O sistema deve oferecer a opção de seleção de comandos separadamente da ativação (navegar pelos menus sem ativá-los).

(A) O sistema deve permitir ao usuário configurar o acesso a opções específicas por meio de comandos como teclas de função (F1 a F12) ou por meio de uma única tecla.

(A) O limite de tempo para execução de comandos no sistema deve ser configurado pelo usuário ou se for o caso de não poder ser configurada, o limite de tempo não deve ser menor que 20 segundos.

(A) O autor deve poder desabilitar opções de flash de conteúdos (imagens) que possam causar foto sensibilidade.

(A) O autor deve poder navegar livremente entre os itens de edição de um conjunto de elementos estruturados (ex: ontologias).

(A) O autor deve poder selecionar conteúdo e executar funções de edição em cada conjunto de elementos estruturados.

(A) O autor deve poder executar pesquisa de texto de todo conteúdo que ele puder editar.

(A) O autor deve poder executar pesquisa de texto de toda marcação que ele puder editar.

(A) O sistema deve permitir que o autor execute opção de desfazer edição.

(A) O sistema deve informar ao autor quais ações do sistema não aceitam a opção de desfazer.

(A) O sistema deve oferecer a opção de salvar e recarregar as configurações relacionadas a saídas visuais ou auditivas e operações do teclado.

(A) O autor deve poder selecionar opções a partir de vários modos de configuração (ex: automática ou personalizada).

(A) O sistema deve oferecer opções de pré-visualização de conteúdo, em conformidade com a necessidade do autor e particularidades do navegador que exibirá o conteúdo final.

(A) O sistema deve obedecer as convenções para configuração de teclado.

(A) As opções do sistema que tenham mesmo termo ou ícone devem desempenhar a mesma função.

(A) Todas as características do sistema que se beneficiam dos requisitos de acessibilidade devem estar disponíveis no help do sistema.

(A) O sistema deve permitir que a configuração de ações disponíveis no sistema sejam disponibilizadas de forma centralizada ou distribuída (lista de opções de tecla, atalhos, nas opções do menu).

(A) A interface de autoria deve ser amigável, de acordo com arquiteturas de plataformas de acessibilidades, ou permitir a interoperabilidade com as mesmas.

(A) O sistema deve permitir a produção de conteúdo acessível a portadores de necessidades especiais.

(A) O sistema deve preservar as opções de acessibilidade durante conversões ou transformações de conteúdos, formatos e plataformas.

(A) Toda marcação de texto e geração de conteúdo automáticas gerada pelo sistema deve estar em conformidade com as recomendações do W3C (WCAG20).

(A) O sistema deve informar e orientar autores a gerar conteúdo acessível de acordo com as opções disponíveis.

(A) O sistema deve conferir e informar aos usuários sobre os problemas de acessibilidade detectados.

(A) O sistema deve permitir que o autor faça a checagem de problemas de acessibilidade no ambiente.

(A) O sistema deve orientar os usuários quanto à solução de problemas de acessibilidade detectados.

(A) O sistema deve oferecer e integrar soluções de acessibilidade.

(A) O sistema deve permitir que as opções de acessibilidade sejam configuráveis.

## ***J.5 Requisitos para ambiente independente de dispositivo***

Nesta seção apresentamos requisitos específicos para ambientes de autoria independente de dispositivo tecnológico (*device independent*), como hardware ou software, plataforma tecnológica, rede de comunicação, etc. com base nas seguintes recomendações do W3C:

- (W3C, 2003) Desafios de autoria para ambientes independentes de dispositivo (*Authoring challenges for device independence*);
- (W3C, 2006) Linguagem de autoria para ambiente independente de dispositivo (*Device independent authoring language - DIAL*);
- (W3C 2005) Glossário de termos para ambiente independente de dispositivo (*Glossary of terms for device independence*).

(A) Os autores devem poder criar documentos com os mesmos tipos de função disponíveis na Web hoje, e disponibilizá-los em diversos tipos de dispositivos tecnológicos e contextos.

(A) O autor deve poder usufruir de funcionalidades adicionais de novos dispositivos, redes e contextos de publicação de documentos, sem comprometimento de acesso a outros dispositivos que não possuam as mesmas funcionalidades que o ambiente de produção.

(A) O autor deve poder produzir documentos que possam ser acessados por outros dispositivos e redes com outras funcionalidades.

(A) O sistema deve minimizar o esforço de navegação em diferentes esquemas de modo a apresentar os resultados da mesma forma em diferentes contextos.

(A) O sistema deve dar suporte à variação da quantidade de conteúdo desenvolvido e apresentá-lo em diferentes dispositivos com diferentes tipos de contextos, e permitir mudança da organização do ambiente para interação do usuário.

(A) O sistema deve permitir o uso de diferentes versões e tipos de mídias, como imagens e áudio, em diferentes dispositivos e contextos.

(A) O sistema deve oferecer aos autores a possibilidade de abstrair ao máximo a especificação de uma interação no ambiente, para reduzir a necessidade de se especificarem interações extras para cada tipo de dispositivo ou contexto.

(A) O sistema deve oferecer a possibilidade de abstração para entrada de dados, seleção de dados, funções de controle e navegação.

(A) O sistema deve tornar fácil a autoria de documentos em outros dispositivos.

(A) O sistema deve dar suporte ao uso de versões diferentes de conteúdo em diferentes dispositivos com contextos distintos.

(A) O sistema deve dar suporte ao uso de versões diferentes de recursos de mídia em diferentes dispositivos com contextos distintos.

(A) O sistema deve prover a funcionalidade de especificação de recursos alternativos para o autor se familiarizar com o novo dispositivo.

(A) O sistema deve prover a funcionalidade de seleção de um recurso apropriado a partir de alternativas disponíveis, de acordo com as capacidades do dispositivo e os atributos do contexto.

(A) O sistema deve prover a funcionalidade de criação de recursos de mídia apropriados, por meio da conversão de um recurso equivalente e com diferentes propriedades, se necessário.

(A) O sistema deve oferecer ao autor o controle sobre os recursos de mídia utilizados em dispositivos e contextos específicos.

(A) O sistema deve oferecer apoio aos autores no gerenciamento da integração entre os dispositivos.

(A) O sistema deve oferecer apoio aos autores no gerenciamento da integração de documentos que contêm conteúdo que independe do dispositivo.

(A) O sistema deve permitir a inclusão e transformação de conteúdos dinâmicos que incorporem marcação que seja livre de definições que afetem diretamente a experiência do autor.

(A) O sistema deve permitir que os autores especifiquem dinamicamente os recursos de mídia a serem usados no contexto do autor.

(A) O sistema deve permitir que os autores utilizem funções avançadas no cliente, incluindo habilidade de executar códigos.

(A) O sistema deve dar suporte à função de controle de interação entre mídias em diferentes dispositivos e contextos.

(A) O sistema deve dar suporte a todo tipo de aplicação, da mais simples à mais complexa.

(A) O sistema deve poder agregar funções e operações para permitir que os elementos da autoria possam ser unidos, conforme suas funcionalidades, facilitando o uso por parte do autor.

(A) O sistema deve permitir o reaproveitamento de aplicações que serão reutilizadas quando da criação de novas aplicações independentes de dispositivos.

(A) O sistema deve permitir que os autores criem aplicações que dêem suporte às diversas funcionalidades do dispositivo.

(A) O sistema deve oferecer mecanismos para que o autor possa comunicar suas necessidades específicas, para poder utilizar as funcionalidades do ambiente.

(A) O sistema deve oferecer mecanismos para que o autor possa definir o layout do documento, o qual pode variar quando utilizado em diferentes dispositivos ou contextos.

(A) O sistema deve guiar o autor durante o período de adaptação ao novo dispositivo ou contexto em uso.

(A) O sistema deve considerar a modalidade do dispositivo durante o período de adaptação do autor.

(A) O sistema deve permitir que autores utilizem diferentes dispositivos com diversas funcionalidades sem a necessidade de conhecimento aprofundado das características de cada dispositivo.

(A) O sistema deve permitir que o autor usufrua das funcionalidades do ambiente sem se sobrecarregar.

(A) O sistema deve permitir que o autor adapte seus conhecimentos e capacidades às realidades de dispositivos específicos.

(A) O sistema deve permitir a reutilização de material gerado pelo autor em dispositivos e contextos diversos.

(A) O sistema deve orientar o autor na reutilização de material gerado, identificando claramente quais materiais são ou não independentes de dispositivo.

### ***K. Atributos do software***

- Interoperabilidade: o ambiente deve permitir a integração com as tecnologias propostas, assim como permitir a viabilidade de integração com outras que possam agregar valor à solução.
- Suporte documentário: os usuários devem poder dispor de documentação como manuais e help online para dúvidas e resolução de problemas.
- Facilidade de uso: o sistema deve apresentar interfaces orientadas a objeto e opções de fácil acesso e utilização para usuários em geral.
- Atratividade: o ambiente deve ser atraente o suficiente aos usuários, preferencialmente com familiaridade com os ambientes correntemente utilizados pelos usuários.
- Acessibilidade: o ambiente deve ser acessível a pessoas com necessidades especiais.
- Adequabilidade: o ambiente deve estar adequado às necessidades específicas dos usuários, considerando-se o contexto em que se dá a produção do documento.
- Confiança: fatores requeridos para estabelecer a confiança no software quando do uso.
- Disponibilidade: fatores que definem o nível de disponibilidade do sistema com pontos de checagem, recuperação, reinicialização.

- Segurança – o sistema deve apresentar controle de acesso, permitir a configuração do ambiente conforme com a política de segurança do contexto em que se usa, deve manter log ou histórico de acessos e de operações executadas.
- Manutenibilidade – neste item são especificados os atributos do software relacionados à facilidade de manutenção da solução.

### ***L. Organizando os requisitos específicos***

De acordo com a norma IEEE Std 830-1998, os requisitos de sistema devem ser organizados de maneira criteriosa, uma vez que normalmente é gerada uma lista muito extensa de requisitos. A norma recomenda 7 modos de organização, com respectivos modelos de uso, são eles:

1. por modo de operação do sistema;
2. por classe de usuários;
3. por objetos;
4. por característica;
5. por estímulo;
6. por resposta;
7. por funções hierárquicas.

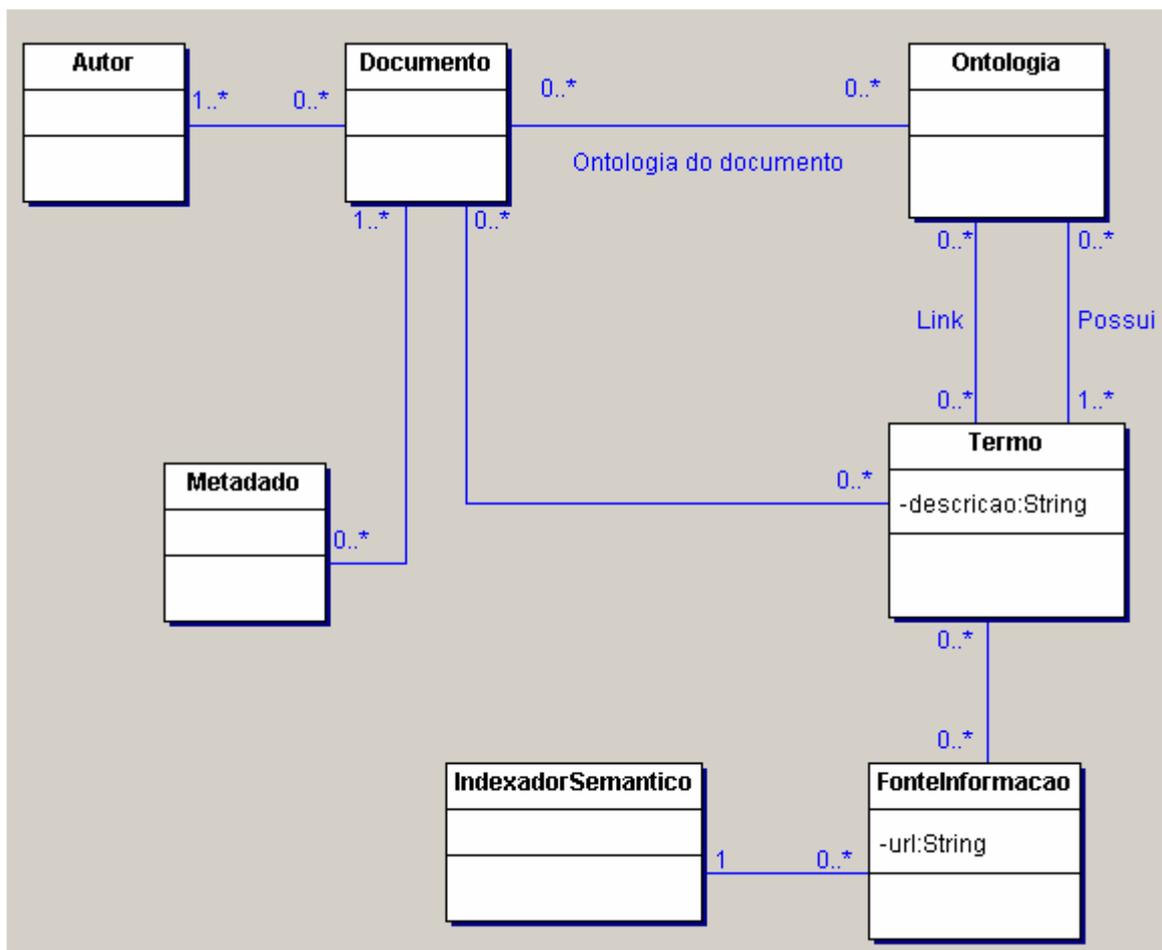
Para a presente especificação de requisitos, optamos por um modo híbrido de organização por modo de operação do sistema e por funções hierárquicas, por serem adequados ao trabalho realizado. Neste modelo é permitida a organização por funcionalidades do sistema, associadas aos respectivos módulos em que se dá a execução das funções pelos atores do sistema. A organização por funções hierárquicas está refletida no item J. Requisitos específicos, deste capítulo. Os requisitos apresentados foram organizados e representados por meio dos diagramas de UML, desenhados a seguir no Anexo III.

## Anexo III Diagramas do Sistema em UML

### III.1 Modelo de Domínio

Elementos do modelo de domínio:

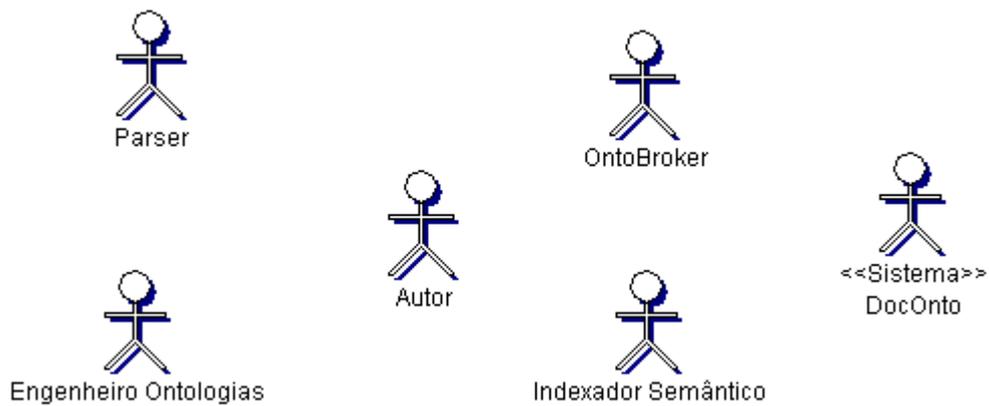
- Autor
- Documento
- Metadado
- Ontologia
- Termo
- Fonte de informação
- Indexador semântico

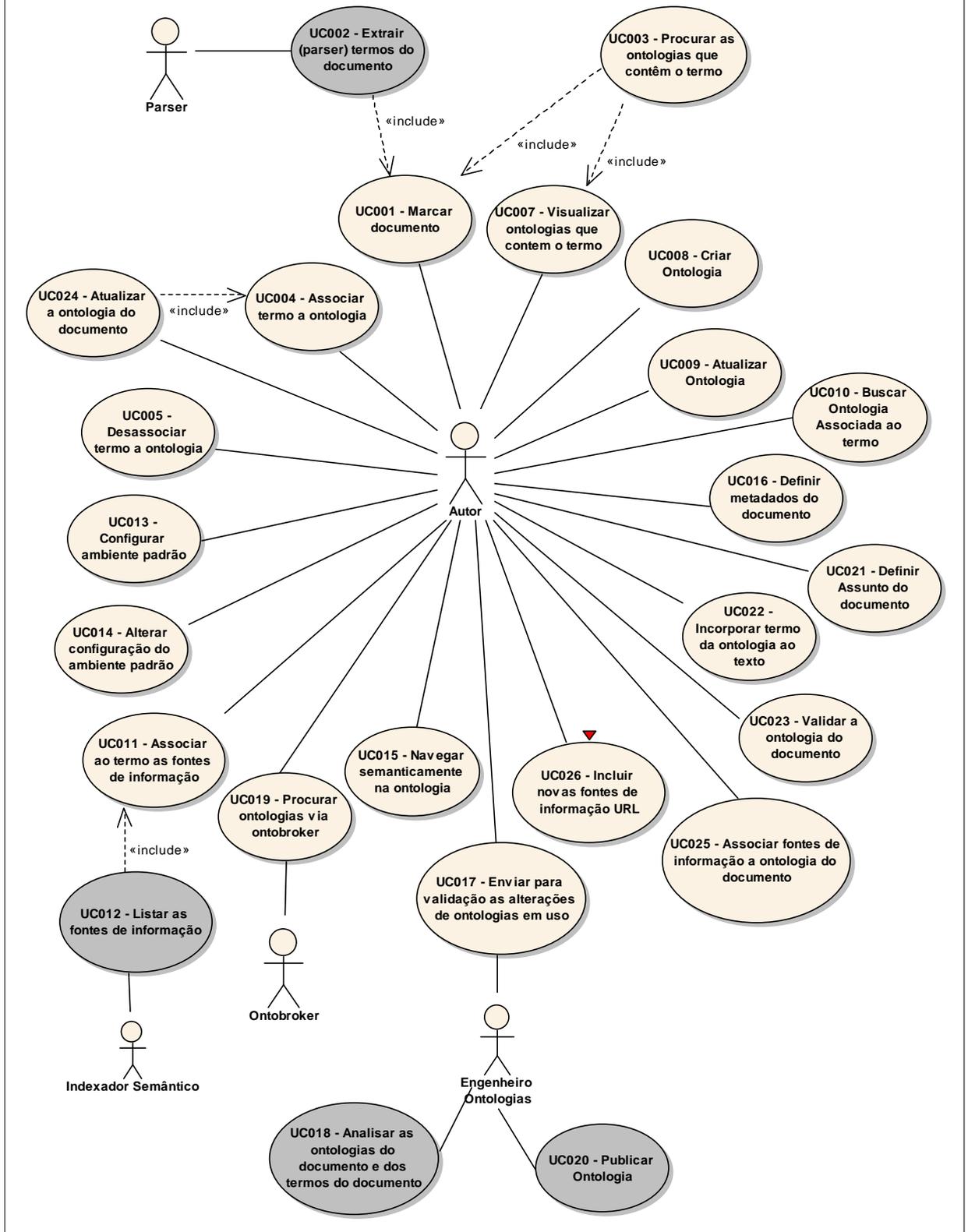


## III.2 Diagrama de caso de uso do sistema

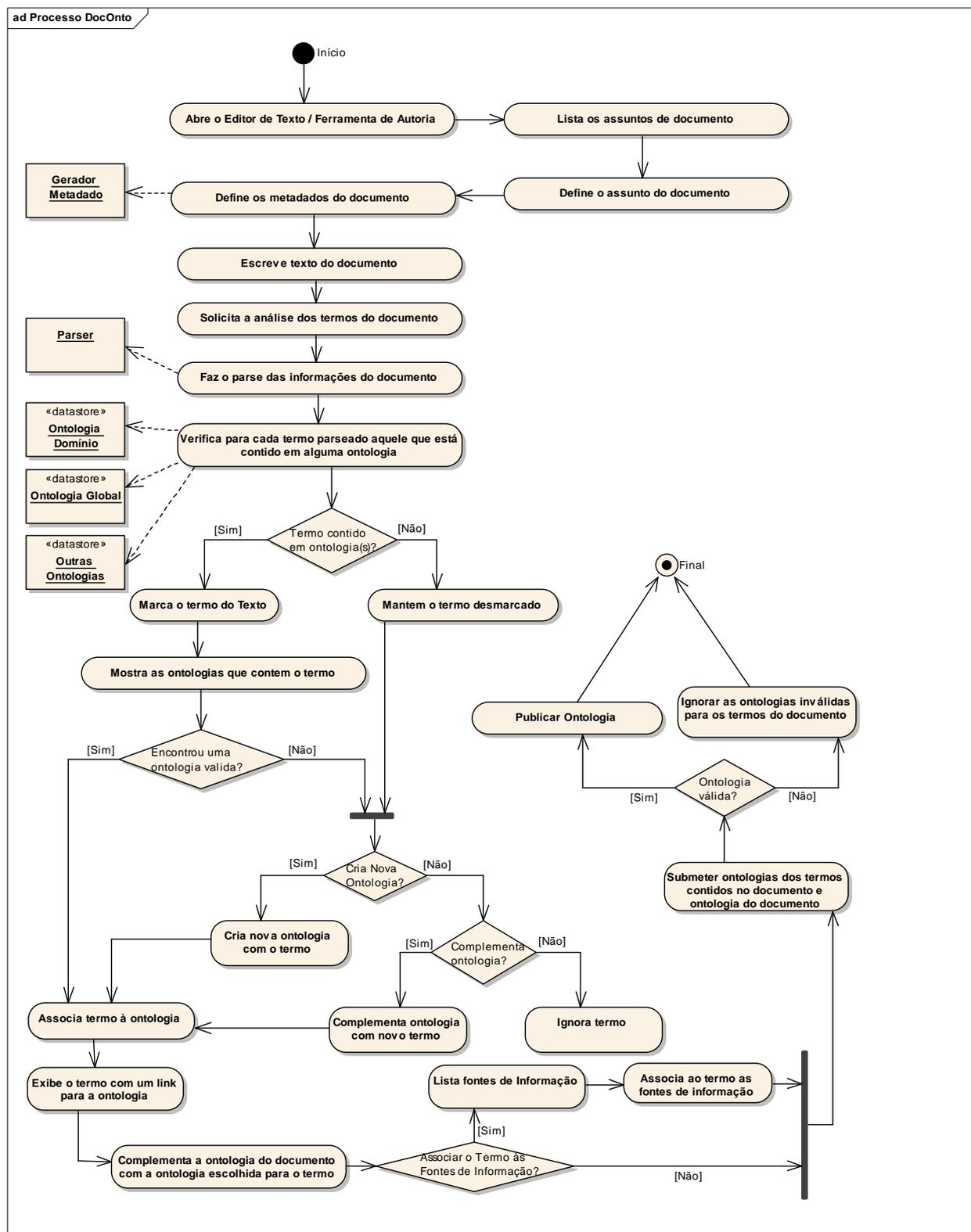
Atores do sistema:

- Autor
- Engenheiro de ontologias
- Ontobroker
- Parser/Extrator
- Indexador semântico
- Sistema DocOnto





### III.3 Diagrama de atividades do sistema

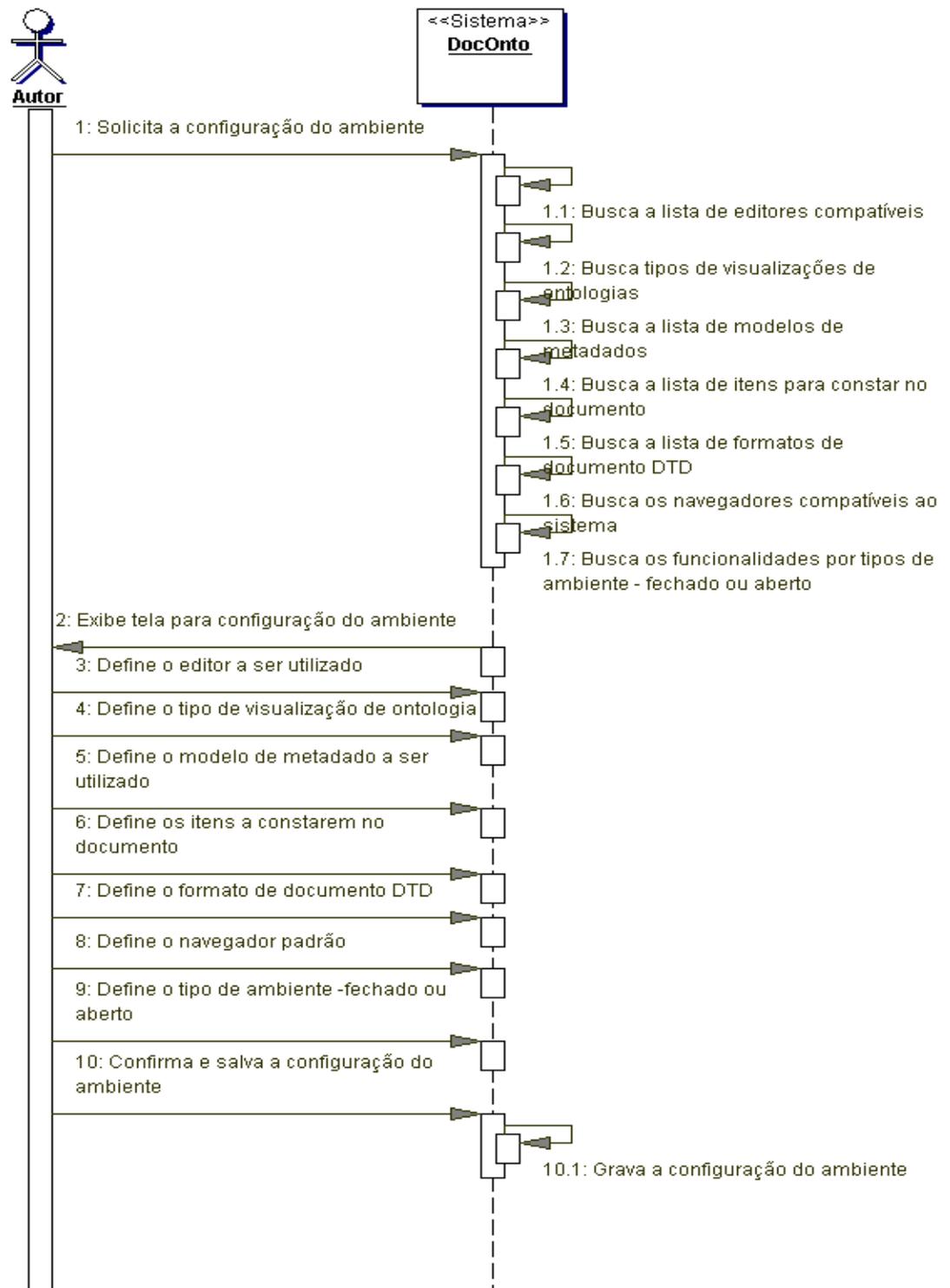


## Descrição geral de ações do diagrama de atividades do sistema

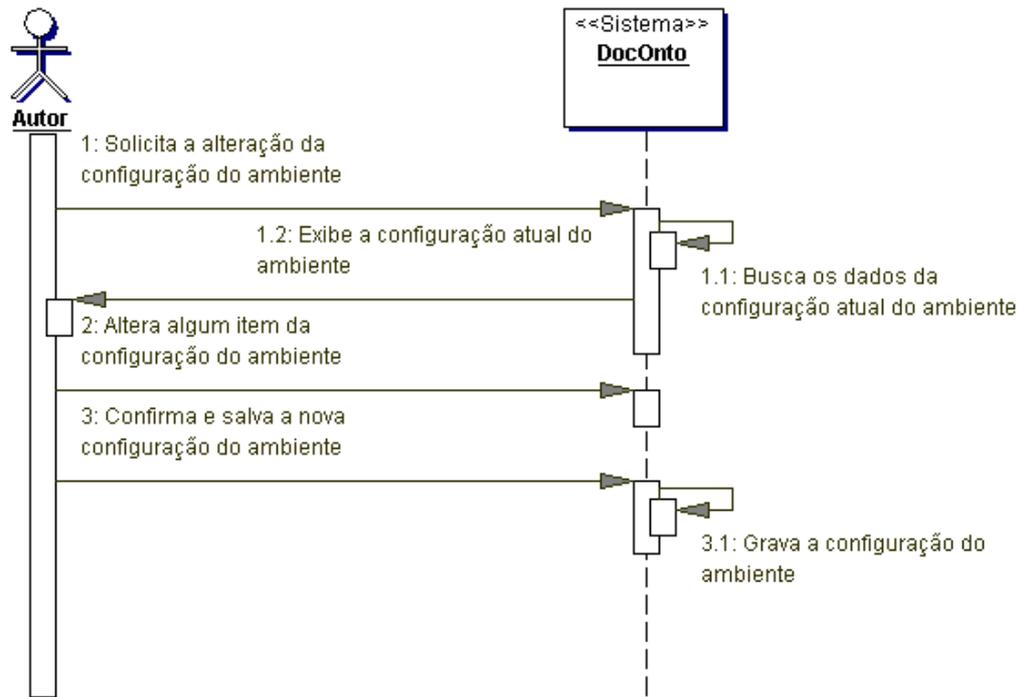
1. O autor pode digitar o texto ou apenas palavras-chave na tela do editor.
2. O extrator (*parser*) pode reconhecer os termos digitados e buscá-los na ontologia. Ao extrair o termo, anota-o e marca-o em tela para visualização e validação do autor. Ao validar, é gerado um *link* entre o termo e a ontologia.
3. Caso os termos não existam ainda, é apresentado na ontologia o conceito que mais se aproxima ao termo digitado (semântica ou sintaticamente). Se houver o conceito ou equivalente na ontologia (via pesquisa do autor), o mesmo poderá validar e salvar a operação.
4. Se o conceito ou equivalente não estiver na ontologia, o autor pode criar a instância do conceito na ontologia, inserir o termo, editar as relações com os termos existentes ou apenas manter o termo novo isolado. Essa ação de inserção será enviada ao BD de ontologias para validação pelo engenheiro de ontologias.
5. O onto-broker poderá buscar em outras ontologias que contém o conceito. Caso não haja ontologia equivalente, pode solicitar a criação da nova ontologia específica.
6. Os termos reconhecidos na ontologia podem ser anotados e marcados como *hiperlink* para a ontologia.
7. O autor pode associar metadados aos termos principais.
8. O documento pode ser representado por meio de uma ontologia específica do documento.
9. A ontologia do documento pode tornar-se uma estrutura de *links* com fontes de informação correlata.
10. O sistema pode sugerir um conjunto de fontes de informação mapeadas para associação a ambiente de discovery.
11. O usuário pode validar a sugestão de fontes de informação.
12. O usuário pode indicar outras fontes de informação – editar endereços (URLs) novos para associação a fontes de conhecimento.
13. O usuário pode submeter o documento para revisão e consequente publicação.

### III.4 Diagramas de seqüência do sistema

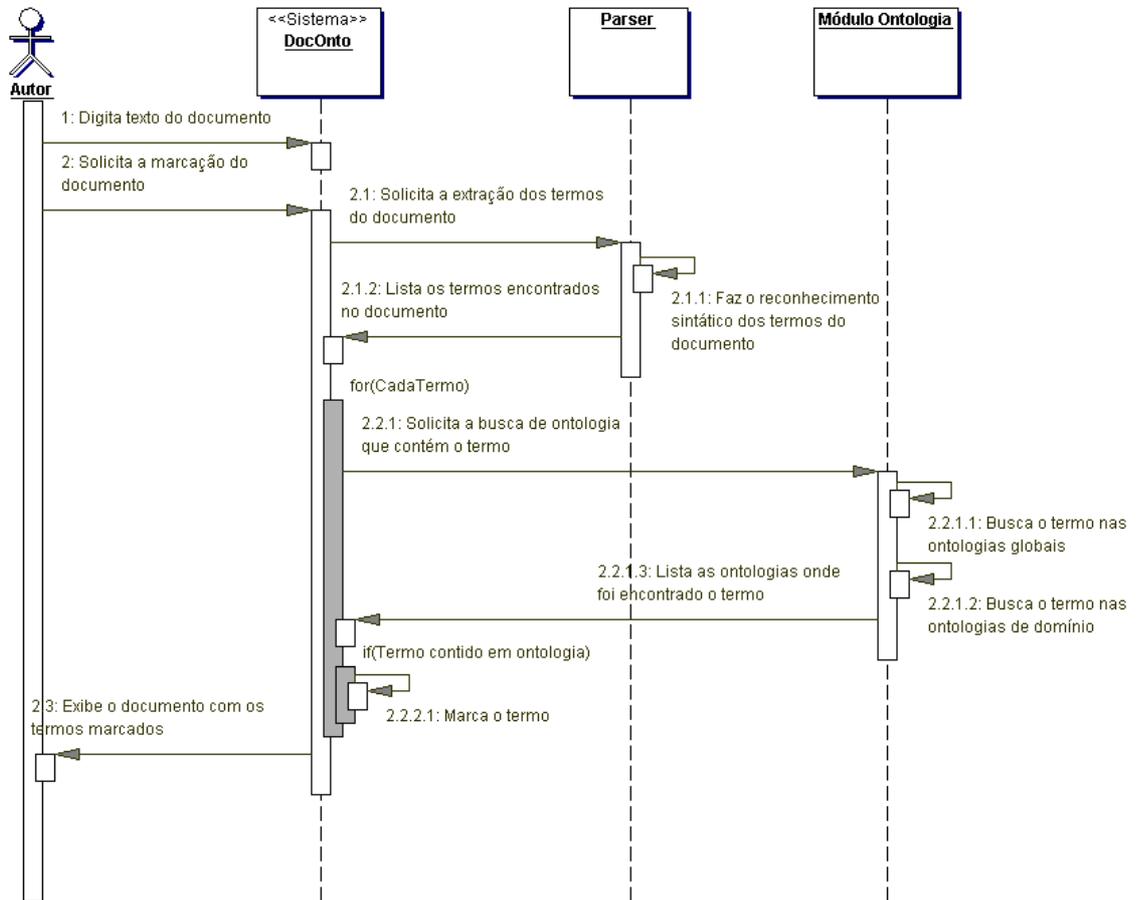
#### i. Configurar ambiente padrão



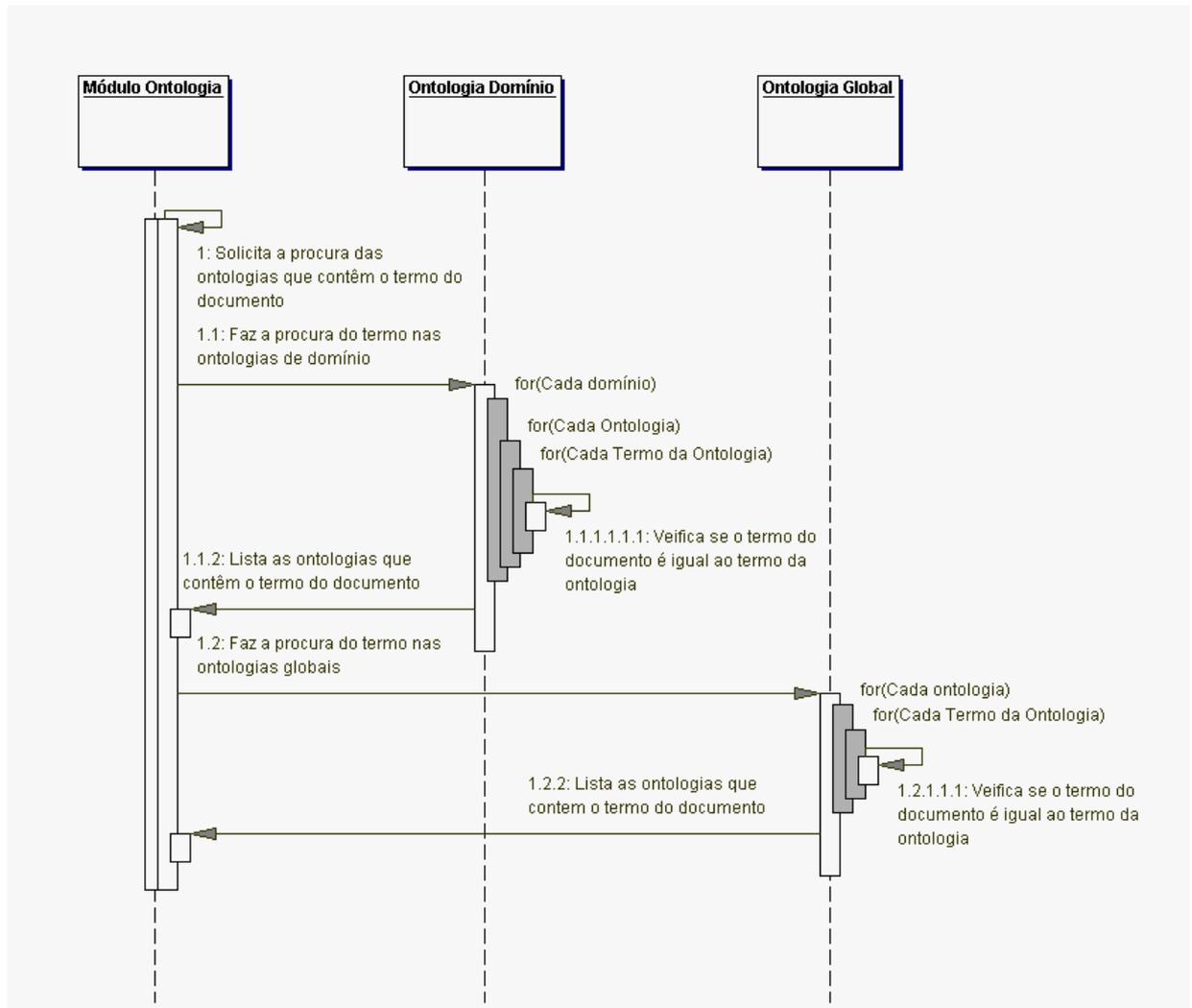
## ii. Alterar configuração do ambiente padrão



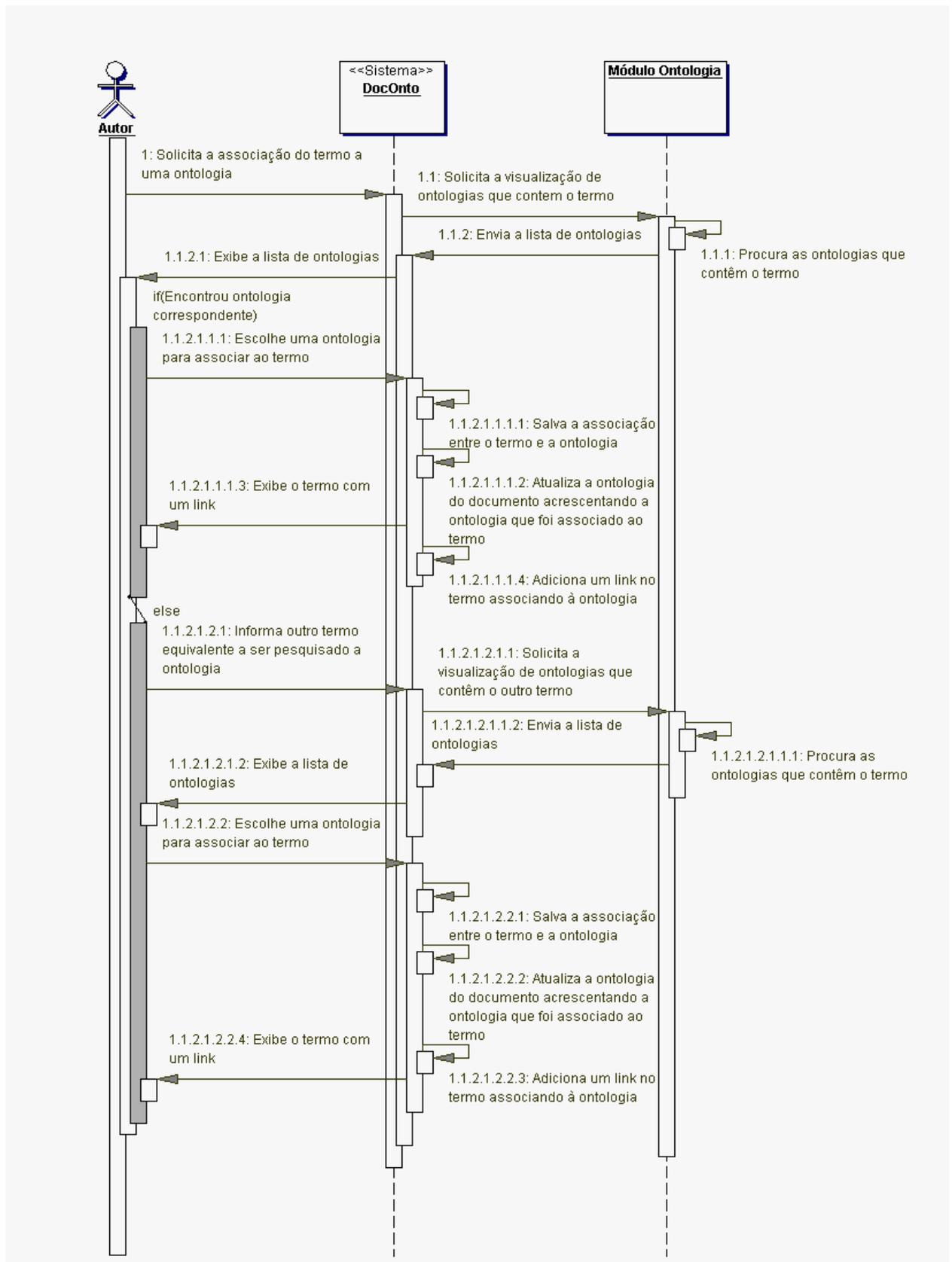
### iii. Marcar documento



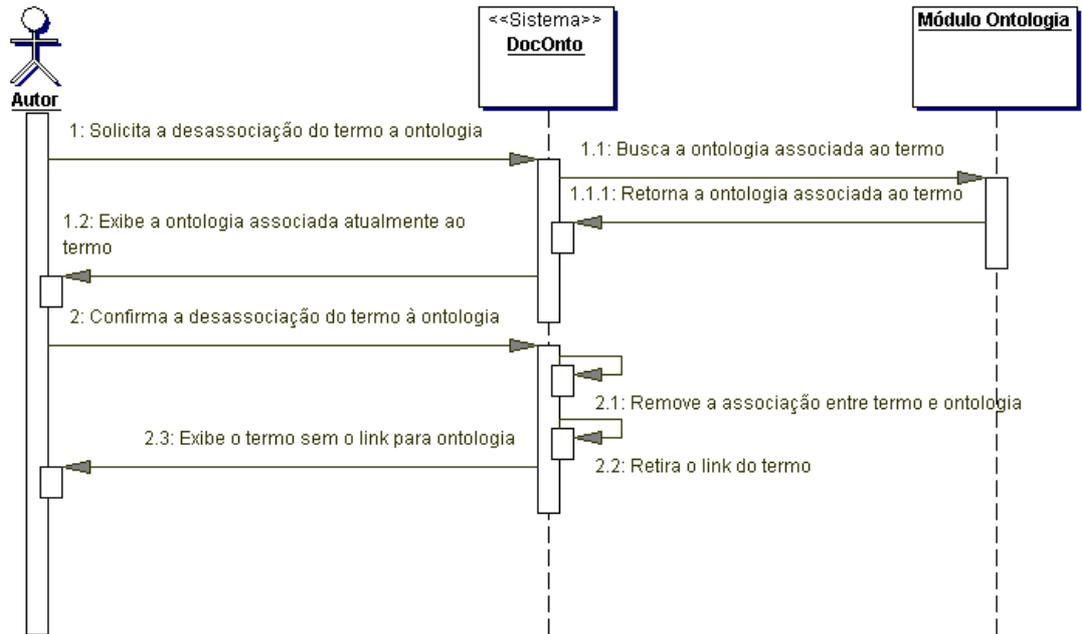
#### iv. Procurar ontologias que contêm o termo



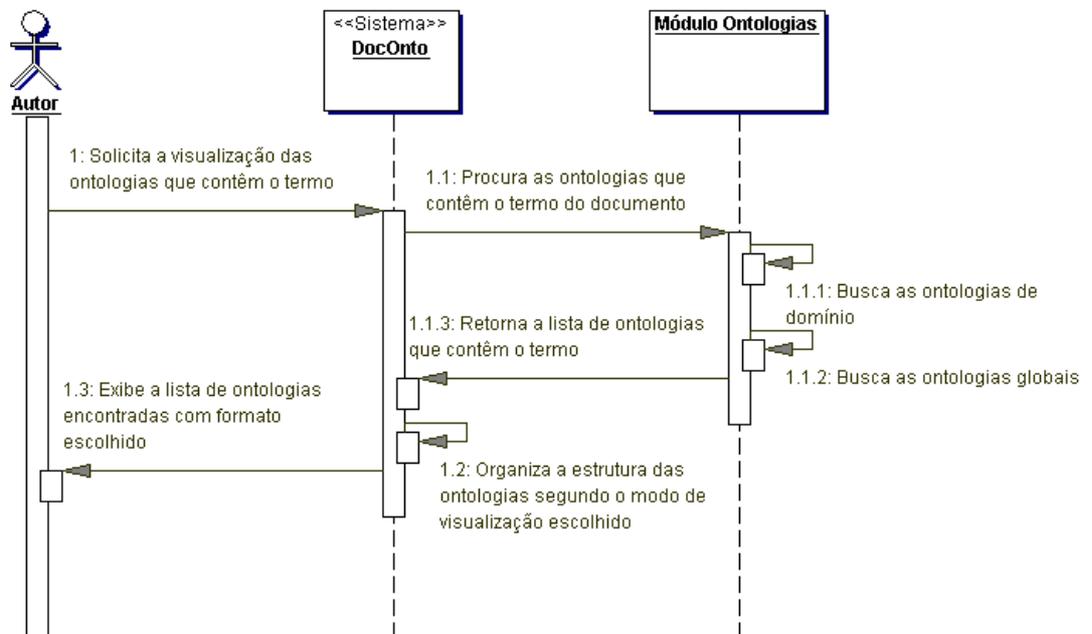
## v. Associar termo a ontologia



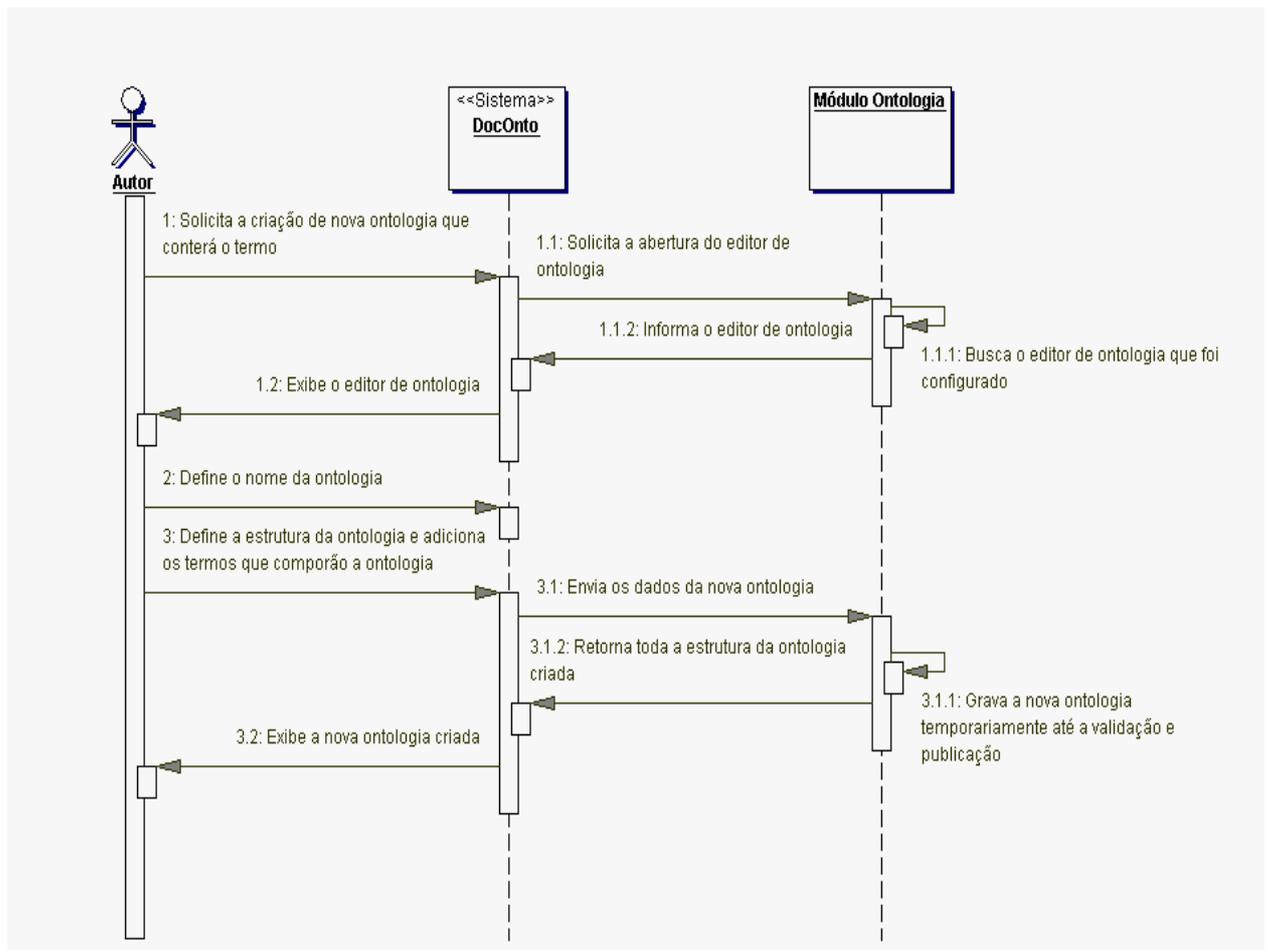
## vi. Desassociar termo da ontologia



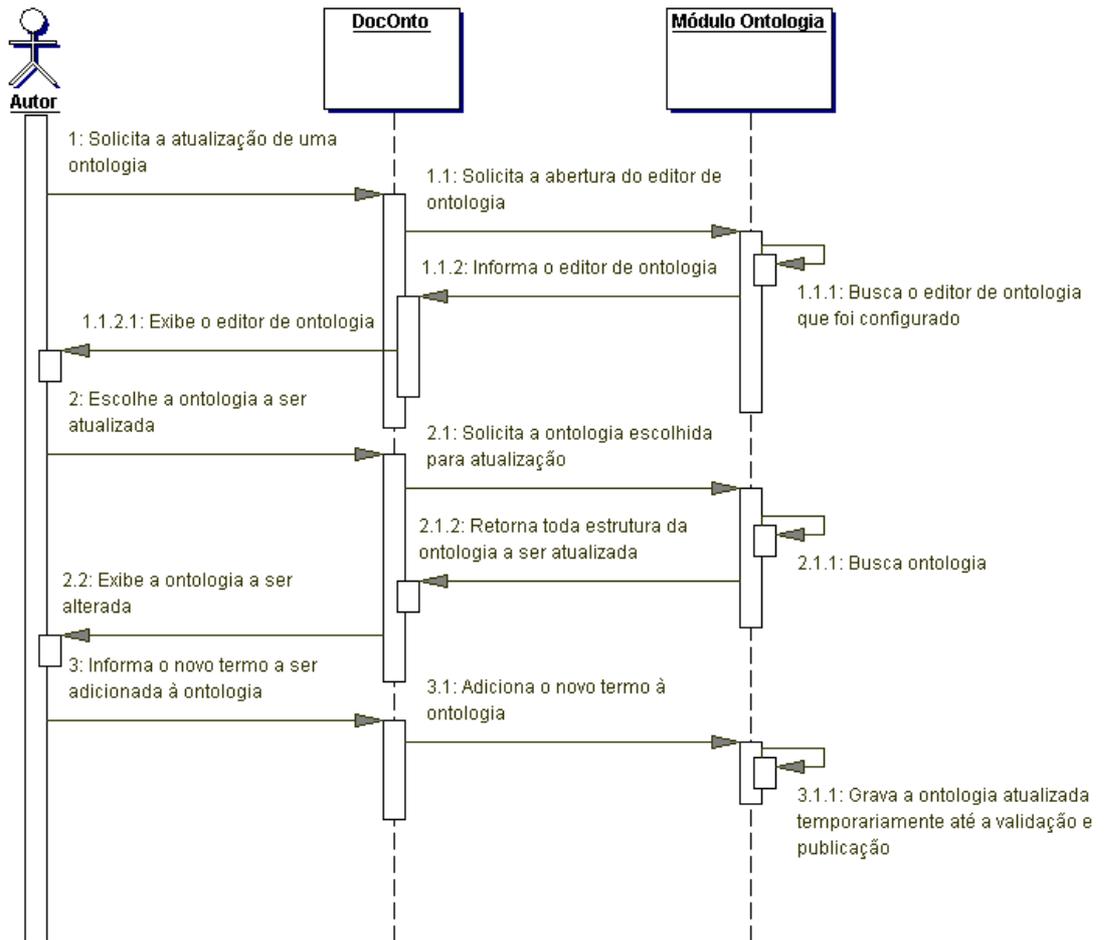
## vii. Visualizar ontologias que contêm o termo



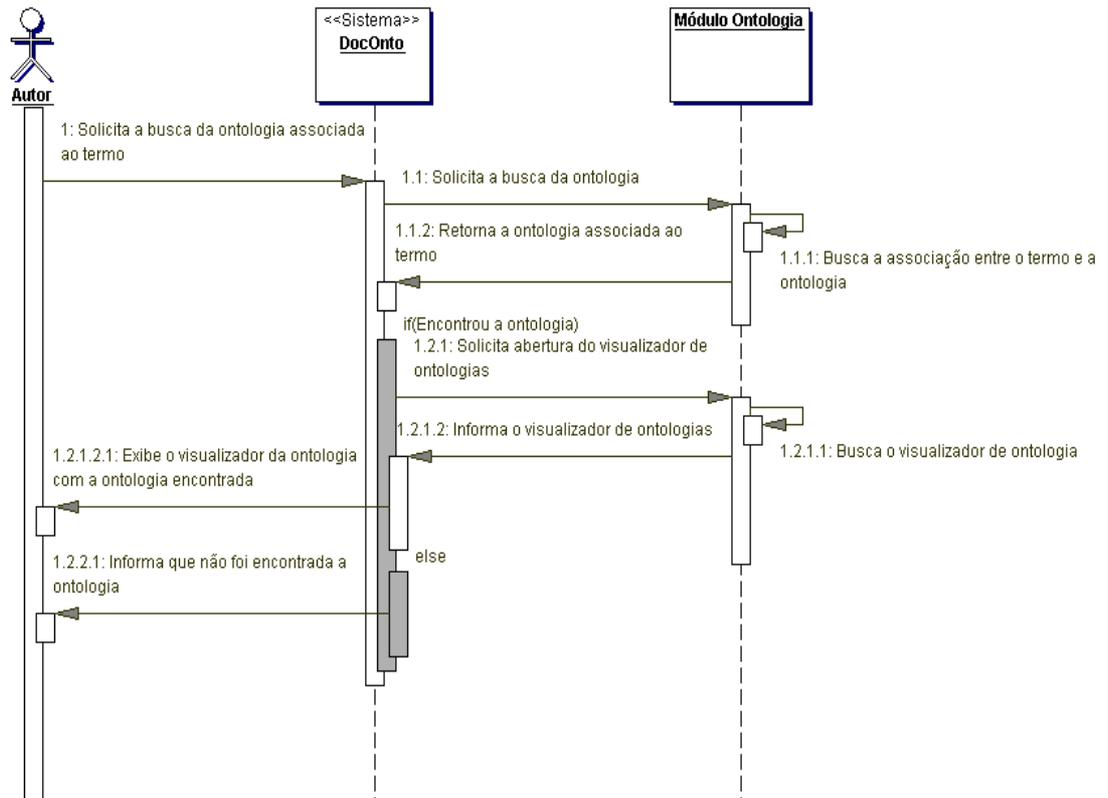
## viii. Criar ontologia



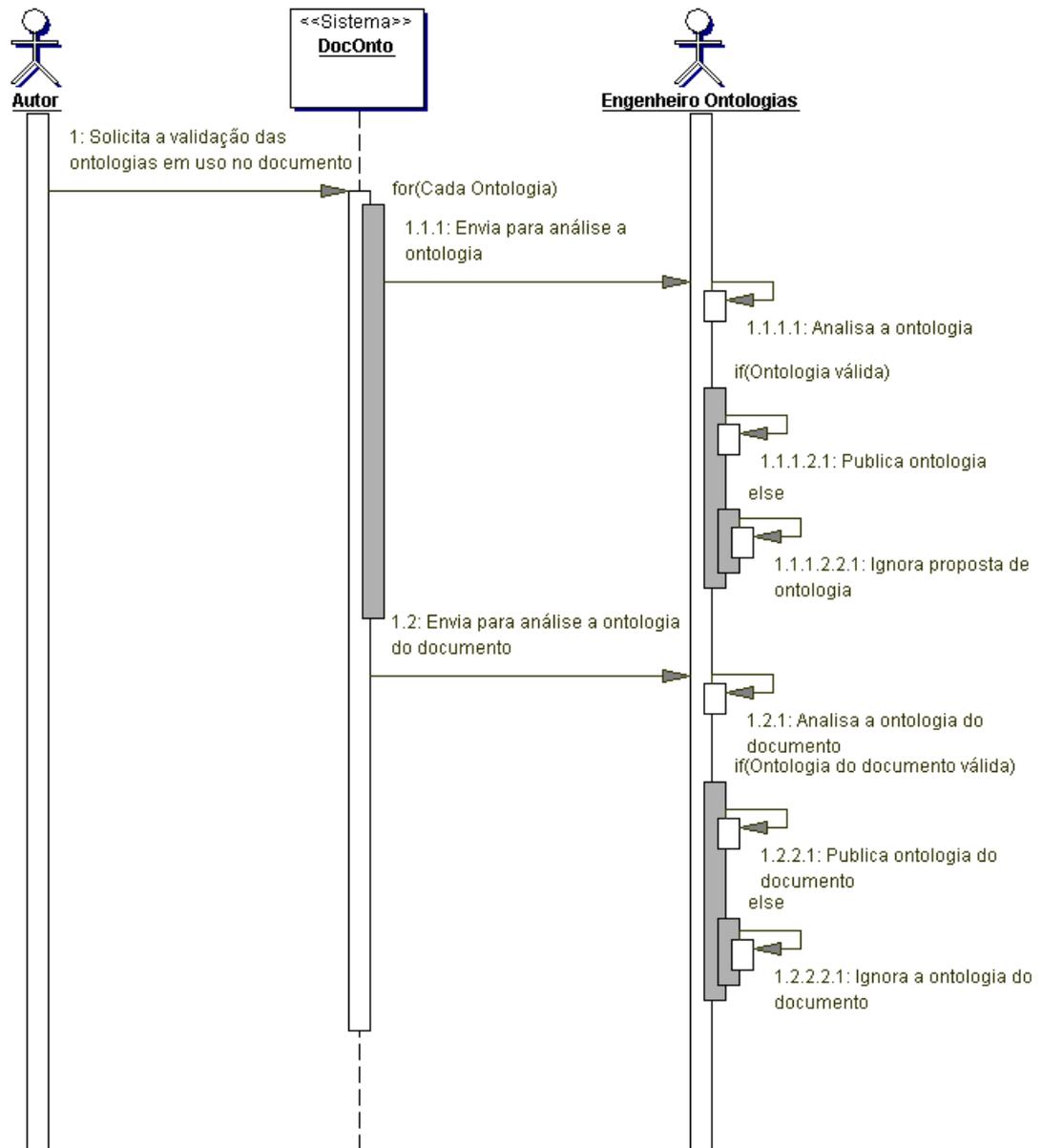
## ix. Atualizar ontologia



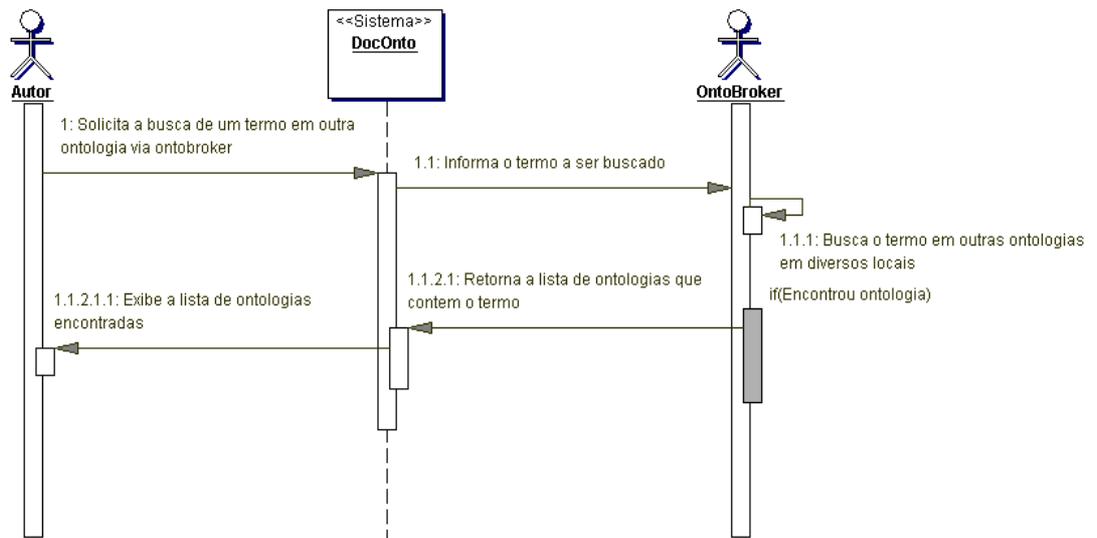
## x. Buscar ontologia associada ao termo



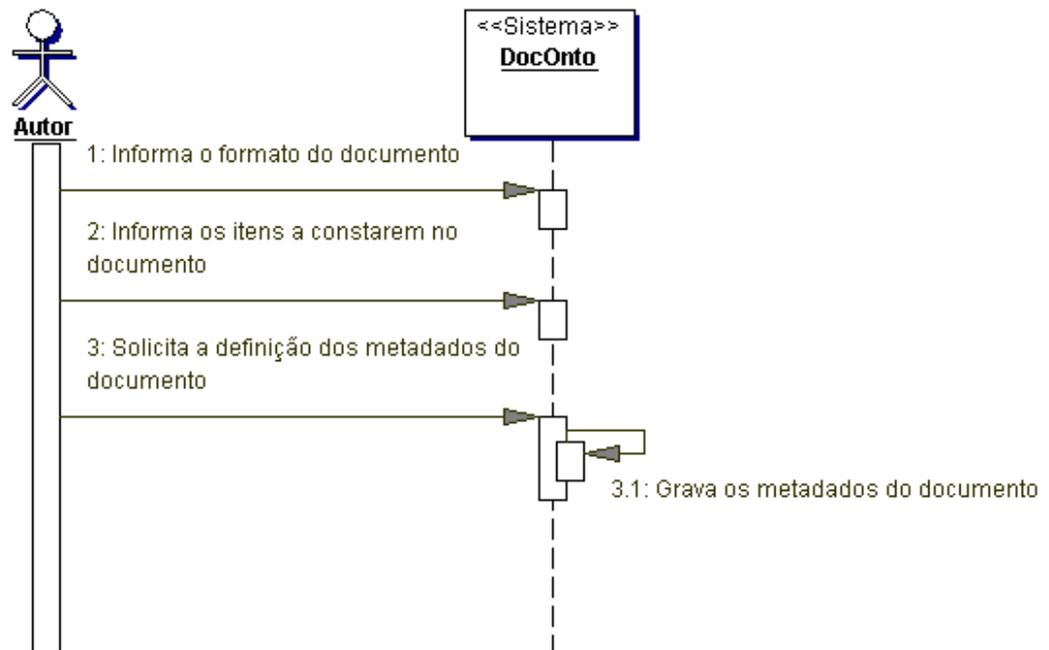
## xi. Enviar para validação as alterações de ontologias em uso



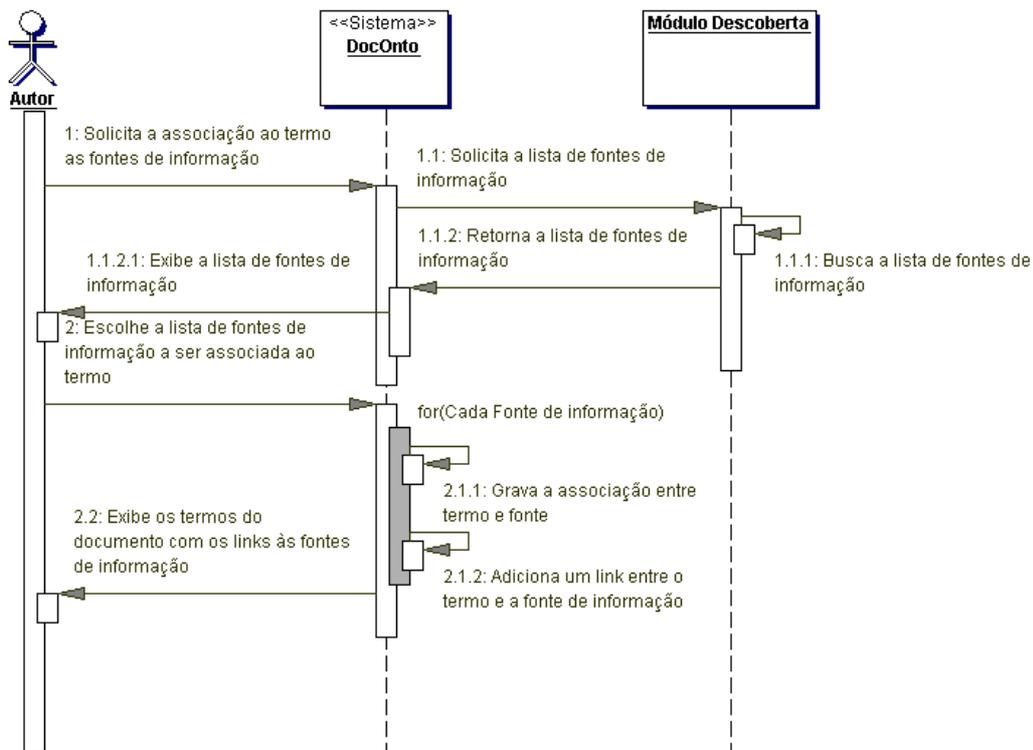
## xii. Procurar ontologias via ontobroker



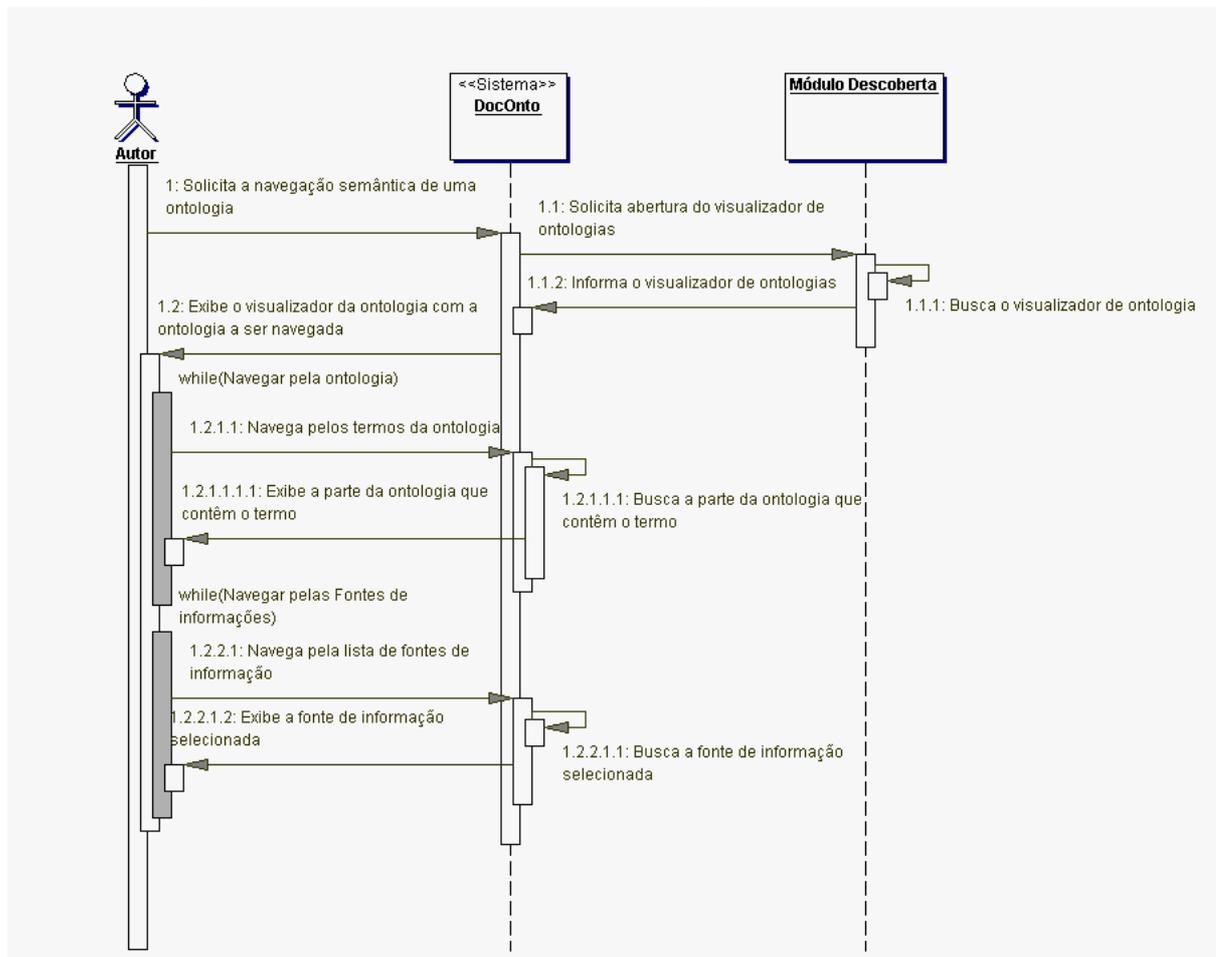
## xiii. Definir metadados do documento



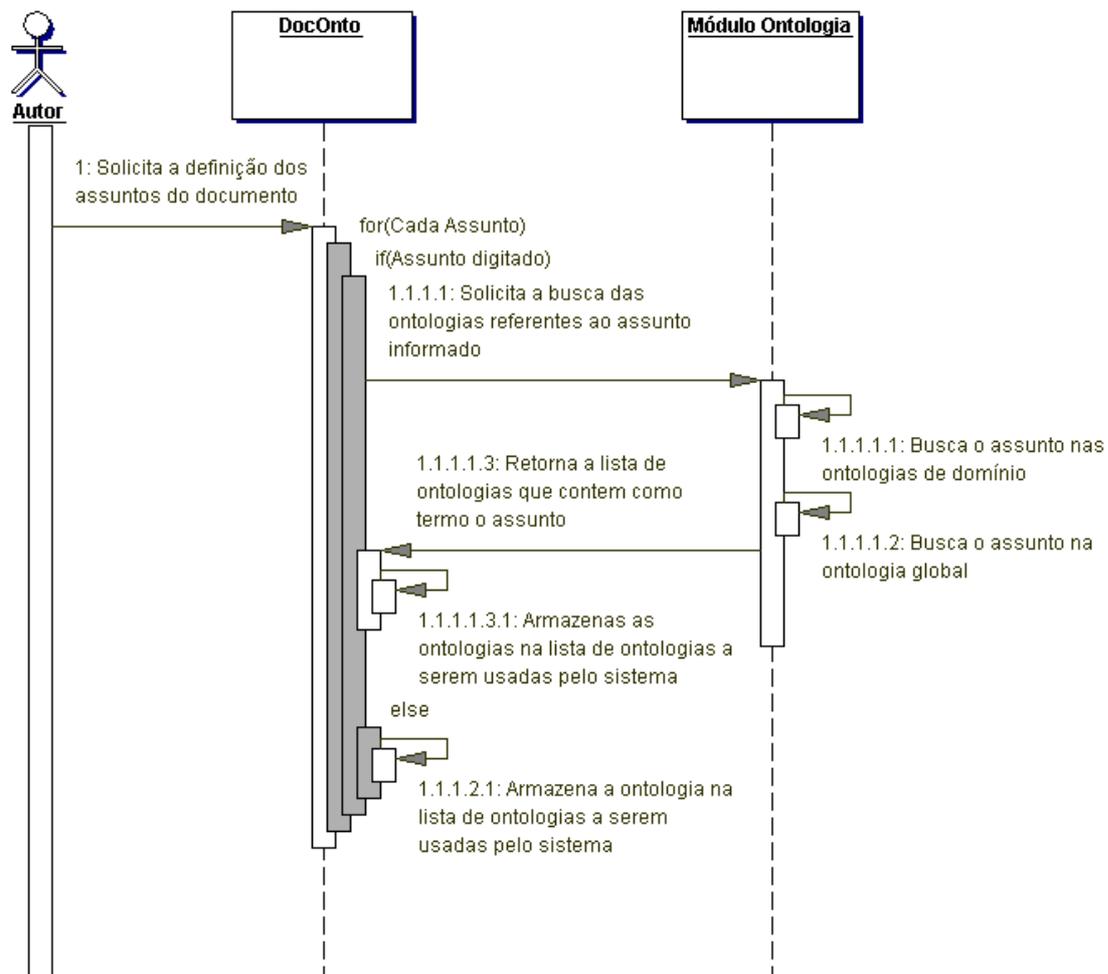
#### xiv. Associar ao termo as fontes de informação



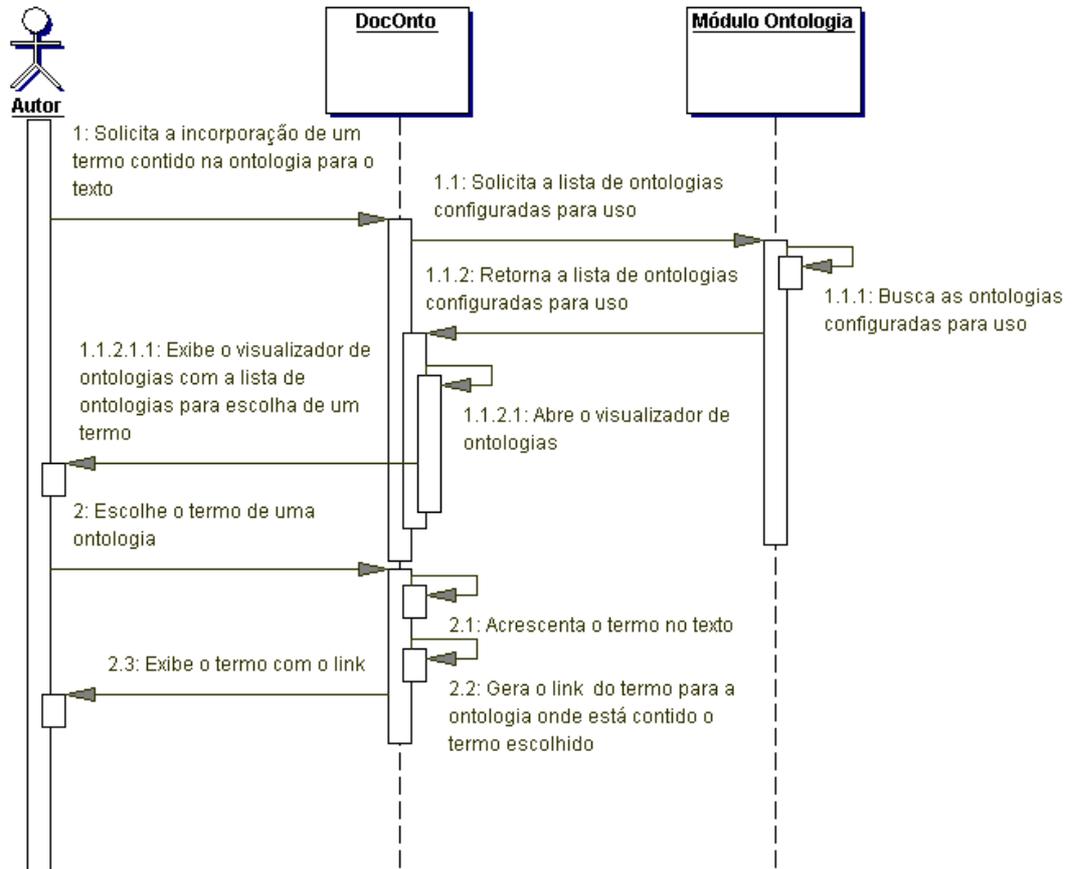
## xv. Navegar semanticamente na ontologia



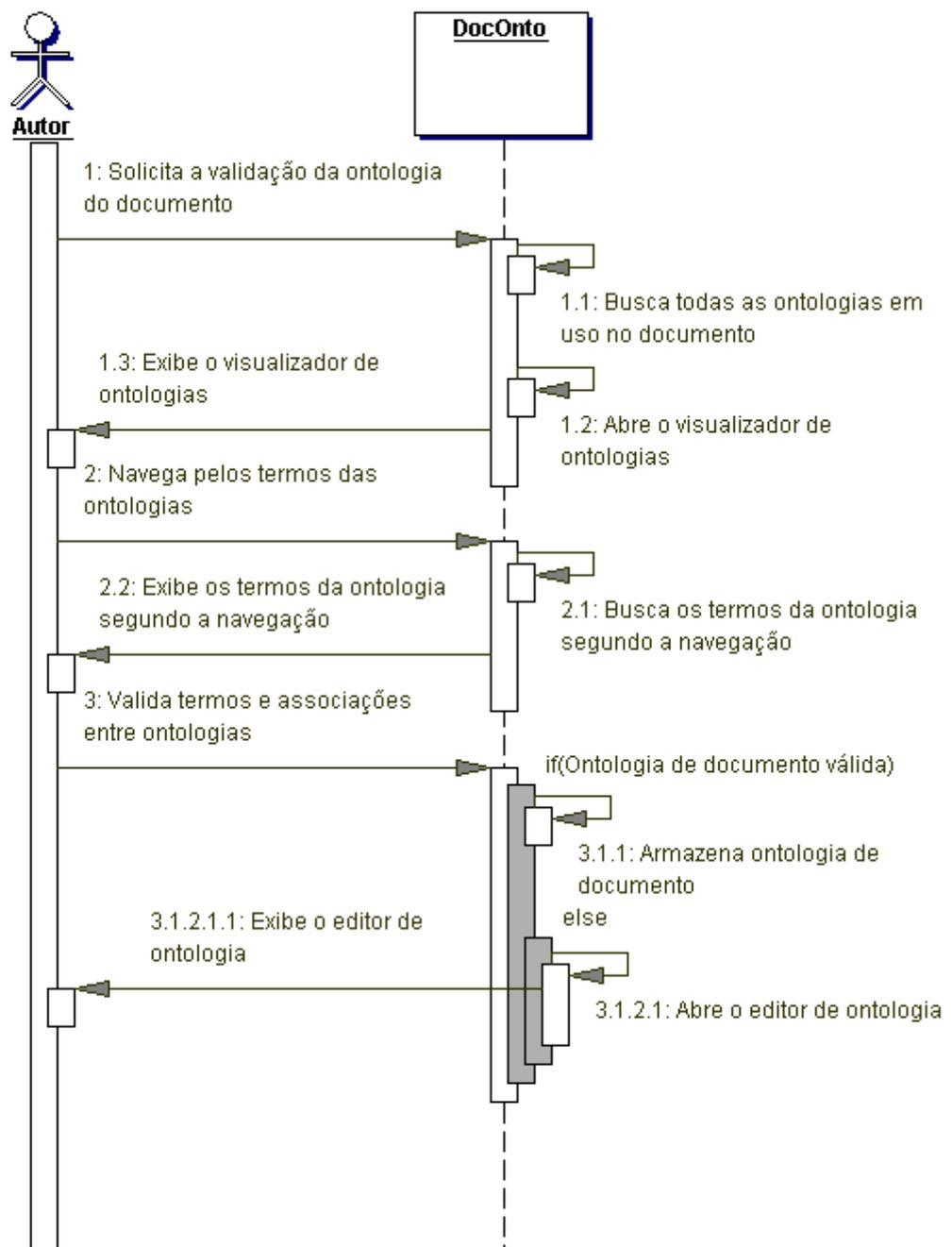
## xvi. Definir assunto do documento



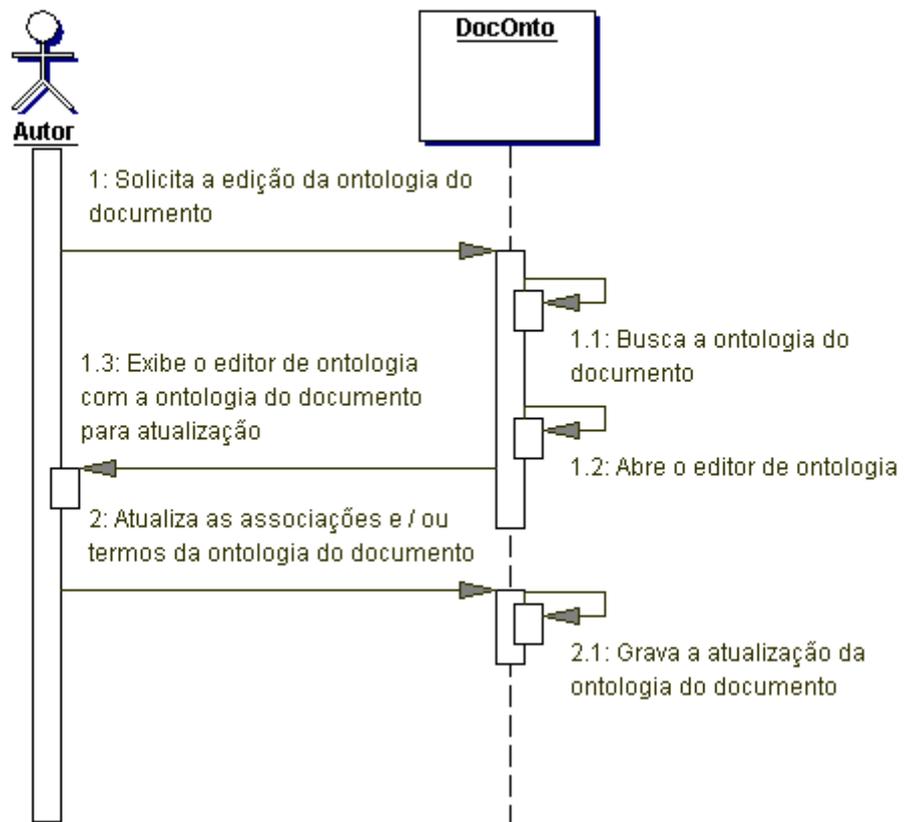
## xvii. Incorporar termo da ontologia ao texto



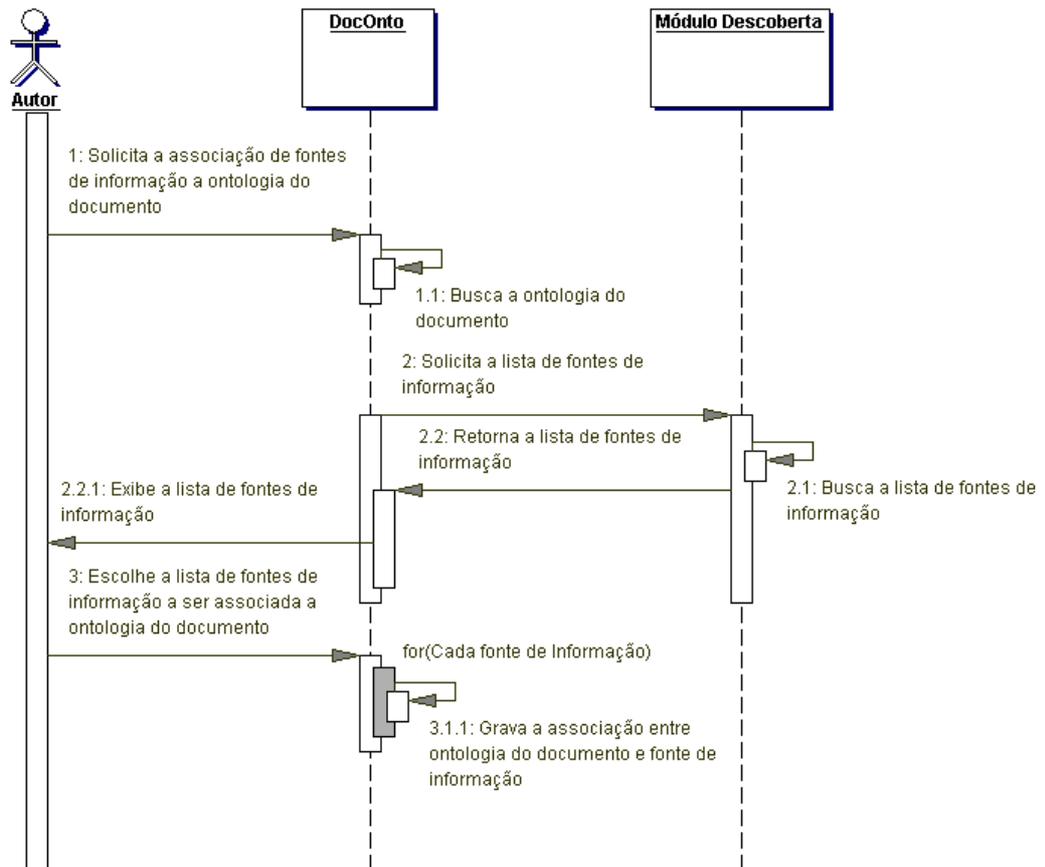
xviii. Validar a ontologia do documento



**xix. Atualizar a ontologia do documento**



**xx. Associar fontes de informação à ontologia do documento**



**xxi. Incluir novas fontes de informação (URL)**

