

**MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
CURSO DE MESTRADO EM SISTEMAS E COMPUTAÇÃO**

CAPITÃO-TENENTE (T) HERMINIO CAMARGO DE SOUZA JUNIOR

**ONTOLOGIAS EMERGENTES: UMA ABORDAGEM PARA CONSTRUÇÃO DE
ONTOLOGIAS A PARTIR DE MAPEAMENTOS PONTO-A-PONTO**

Rio de Janeiro

2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

INSTITUTO MILITAR DE ENGENHARIA

CAPITÃO-TENENTE (T) HERMINIO CAMARGO DE SOUZA JUNIOR

**ONTOLOGIAS EMERGENTES: UMA ABORDAGEM PARA
CONSTRUÇÃO DE ONTOLOGIAS A PARTIR DE MAPEAMENTOS
PONTO-A-PONTO**

Dissertação de Mestrado apresentada ao Curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Sistemas e Computação.

Orientadora: Prof^a Maria Cláudia R. Cavalcanti - D.Sc.

Co-orientadora: Prof^a Ana Maria de C. Moura - D.Eng.

Rio de Janeiro

2008

C2008

INSTITUTO MILITAR DE ENGENHARIA

Praça General Tibúrcio, 80 – Praia Vermelha

Rio de Janeiro - RJ CEP: 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do autor e do(s) orientador(es).

004.678 S729	Souza Junior, Herminio Camargo Ontologias Emergentes: Uma abordagem para construção de ontologias a partir de mapeamentos ponto-a-ponto , Capitão-Tenente(T) Herminio Camargo de Souza Junior. - Rio de Janeiro: Instituto Militar de Engenharia, 2008. 151p.: 92 il., 3 tab. Dissertação: (mestrado) - Instituto Militar de Engenharia, Rio de Janeiro, 2008. 1. Web Semântica - ontologias. 2. Sistemas P2P - integração. I. Instituto Militar de Engenharia. II. Título. CDD 004.678
-----------------	---

INSTITUTO MILITAR DE ENGENHARIA

CAPITÃO-TENENTE (T) HERMINIO CAMARGO DE SOUZA JUNIOR

**ONTOLOGIAS EMERGENTES: UMA ABORDAGEM PARA
CONSTRUÇÃO DE ONTOLOGIAS A PARTIR DE MAPEAMENTOS
PONTO-A-PONTO**

Dissertação de Mestrado apresentada ao Curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Sistemas e Computação.

Orientadora: Prof^a Maria Cláudia Reis Cavalcanti - D.Sc.

Co-orientadora: Prof^a Ana Maria de Carvalho Moura - Dr. Ing.

Aprovada em 23 de junho de 2008 pela seguinte Banca Examinadora:

Prof^a. Maria Cláudia Reis Cavalcanti - D.Sc. do IME - Presidente

Prof^a. Ana Maria de Carvalho Moura - Dr. Ing. do IME

Prof^a. Claudia Marcela Justel - D.Sc. do IME

Prof. Geraldo Bonorino Xexéo - D.Sc. da DCC-UFRJ

Rio de Janeiro

2008

AGRADECIMENTOS

Agradeço a Deus pela graça alcançada.

A minha família que muito me encorajou e ajudou. A minha mãe e meu pai que sempre insentivaram a busca de conhecimento e crescimento pessoal, dedicando todo tempo e atenção, até em momento difíceis.

Aos meus amigos que suportaram e compreenderam as minhas ausências.

A minha orientadora professora Yoko e co-orientadora professora Ana Maria por terem me conduzido com muita paciência, amizade e profissionalismo.

A professora Claudia Justel e o professor Xexéo pelas importantes sugestões de melhoria deste trabalho.

Ao Major Miranda muito obrigado pelas explicações do COPPEER.

Agradeço também aos colegas do mestrado em especial a Fernanda que contribuiu muito no relatório técnico e em outros trabalhos.

A todos os professores e colaboradores que direta e indiretamente nortearam este trabalho facilitando o alcance de seus objetivos.

A todos os professores da disciplina de seminários tecnológicos pela disponibilidade e vontade de dividir suas experiências.

A todas as entidades de fomento a educação e pesquisa que ajudam cada vez melhorar o Brasil.

A Marinha do Brasil por confiar e acreditar na possibilidade de contribuição que um curso de mestrado pode proporcionar as Organizações Militares.

A Diretoria de Administração da Marinha e a Pagadoria de Pessoal da Marinha por terem conduzido todos esforços para proporcionar a conclusão deste curso.

Aos todos os meus amigos da Marinha pelo incentivo e parceria, em especial a minha amiga Kátia Allevato.

A todos aqueles que colaboram para educação e pesquisa no Brasil.

SUMÁRIO

LISTA DE ILUSTRAÇÕES.....	8	
LISTA DE TABELAS.....	12	
1	INTRODUÇÃO	15
1.1	Motivação.....	15
1.2	Objetivo.....	17
1.3	Organização da dissertação.....	17
2	INTEROPERABILIDADE DE ONTOLOGIAS	19
2.1	Ontologias.....	19
2.3	Ferramentas de cálculo de similaridades entre ontologias.....	24
2.4	Tecnologias para interoperabilidade entre ontologias.....	24
2.4.1	Tecnologias para Mapeamento.....	25
2.4.1.1	MAFRA (MAEDCHE et al, 2002) e (Nuno Silva, 2004).....	25
2.4.2	Tecnologias para Alinhamento.....	28
2.4.2.1	CATO (FELICISSIMO, 2004).....	28
2.4.3	Tecnologias para Combinação ou Fusão.....	30
2.4.3.1	ONTOMERGE (DOU et al, 2002).....	30
2.4.4	Tecnologias para Integração.....	31
2.4.4.1	ONTOLINGUA (Farquhar et al, 1996).....	31
2.5	Considerações finais.....	33
3	INTEGRAÇÃO DE DADOS EM AMBIENTES AUTÔNOMOS	34
3.1	Grades.....	35
3.2	Sistemas P2P.....	37
3.2.1	Bibster (Broekstra et al, 2004).....	41
3.2.2	OYSTER (PALMA et al, 2005).....	43
3.2.3	ROSA - P2P (BRITO & MOURA, 2005).....	45
3.2.4	XPEER (BELLAHSÈNE et al., 2004).....	47
3.2.5	SPEED (DAMIRES, 2007) E (PIRES, 2007).....	49
3.3	Grade x P2P.....	50
3.4	Considerações finais.....	53

4	ONTOLOGIAS EMERGENTES	54
4.1	Arquitetura	55
4.1.1	Dinâmica de mapeamentos e geração da OE	56
4.1.2	Módulo de Mapeamento P2P	59
4.1.3	Módulo de Gerência	60
4.1.4	Módulo de Geração da OE	60
4.1.5	Módulo de Navegação-Instanciação	61
4.1.6	Módulo de rastreamento	63
4.1.7	Bases de dados	64
4.2	Definições de Heurísticas	65
4.2.1	Definições básicas	65
4.2.2	Definições de constantes	74
4.2.3	Heurísticas para composição da OE	75
4.3	Considerações finais	77
5	PROTÓTIPO DO SISTEMA	79
5.1	Visão geral do protótipo	79
5.2	Implementação a partir dos módulos da arquitetura (diagramas)	80
5.2.1	Módulo de Mapeamento Ponto-a-Ponto	80
5.2.2	Módulo de gerência	82
5.2.3	Módulo de Geração da OE	86
5.2.4	Módulo de Comunicação e Gerência Ponto-a-Ponto	96
5.2.4.1	COPPEER - infra-estrutura para gerência P2P (MIRANDA, 2006)	97
5.2.4.2	Implementação do Módulo de Comunicação e Gerência P2P	98
5.3	Considerações finais	99
6	EXEMPLO DE USO E AVALIAÇÃO DO PROTÓTIPO	101
6.1	Exemplo de uso do protótipo OntoEmerge	101
6.1.1	Construção das Ontologias	102
6.1.1.1	Ontologia CPAD	103
6.1.1.2	Ontologia LEGISMAR	105
6.1.1.3	Ontologia SIGDEM	106
6.1.1.4	Ontologia PESSOAL	107
6.1.2	Mapeamentos realizados	108
6.1.3	Análise dos mapeamentos	113

6.1.4	Geração da Ontologia Emergente (OE)	113
6.2	Avaliação do protótipo OntoEmerge.....	122
6.2.1	Descrição das ontologias utilizadas pela bancada de testes	122
6.2.2	Descrição da bancada de testes	128
6.2.3	Experimentos realizados	131
6.2.4	Análise dos resultados obtidos	132
6.3	Considerações finais.....	135
7	CONCLUSÃO	137
7.1	Contribuições	138
7.2	Melhorias e trabalhos futuros	139
7.2.1	Melhorias na proposta	139
7.2.2	Melhorias no protótipo OntoEmerge.....	140
7.2.3	Sugestões para trabalhos futuros	140
8	REFERÊNCIAS BIBLIOGRÁFICAS	142

LISTA DE ILUSTRAÇÕES

FIG. 2.1	Classificação das ontologias (GUARINO, 1998).....	21
FIG. 2.2	Estrutura da SBO (MAEDCHE et al, 2002).....	26
FIG. 2.3	Arquitetura conceitual do MAFRA (MAEDCHE et al, 2002).....	28
FIG. 2.4	Estratégia para alinhamento taxonômico de ontologias (FELICISSIMO, 2004).....	29
FIG. 2.5	Arquitetura do OntoMerge (DOU et al, 2002).....	31
FIG. 2.6	Arquitetura do Ontolingua (FARQUHAR et al, 1996).....	32
FIG. 3.1	Evolução da Grade (DE ROURE et al, 2003).....	36
FIG. 3.2	Arquitetura Pura.....	39
FIG. 3.3	Arquitetura Parcialmente Centralizada.....	39
FIG. 3.4	Arquitetura Híbrida.....	40
FIG. 3.6	Arquitetura do Bibster (BROEKSTRA et al, 2004).....	43
FIG. 3.7	Arquitetura do ROSA-P2P (BRITO, 2005).....	46
FIG. 3.8	Organização da fontes de dados do XPeer (BELLAHSÈNE et al, 2004).....	48
FIG. 3.9	Extensão do XPeer (BELLAHSÈNE et al, 2006).....	48
FIG. 3.10	Arquitetura geral do sistema SPEED (PIRES, 2007).....	50
FIG. 4.1	Arquitetura Ontologias Emergentes.....	56
FIG. 4.2	Exemplo de dinâmica dos mapeamentos entre quatro pontos.....	58
FIG. 4.3	Exemplo de mapeamentos entre as ontologias 1 e 2.....	59
FIG. 4.4	Instâncias das ontologias 1 e 2.....	60
FIG. 4.5	Exemplo utilizado para demonstração do módulo de Navegação-Instanciação.....	62
FIG. 4.6	Visão local do ponto 1.....	63
FIG. 4.7	Exemplo de mapeamentos entre conceitos de Ontologia, que formam um grafo conexo.....	70
FIG. 4.8	Exemplo de mapeamentos entre ontologias envolvendo quatro pontos.....	76
FIG. 5.1	Arquitetura conceitual do OntoEmerge.....	80
FIG. 5.2	Representação da Classe EXPEDIENTE da ontologia SIGDEM em RDF.....	81
FIG. 5.3	Representação da Classe DOCUMENTOS da ontologia CPAD em RDF.....	81
FIG. 5.4	Informação de mapeamento da ontologia SIGDEM para CPAD.....	81

FIG. 5.5	Representação de mapeamentos entre ontologias - SBO estendido.....	82
FIG. 5.6	Grafo dos mapeamentos realizados entre as ontologias.....	83
FIG. 5.7	Arquivo produzido pela interface de configuração.....	85
FIG. 5.8	Exemplo de mapeamentos entre conceitos - ponte semântica (ConceptBridge).....	86
FIG. 5.9	Exemplo de mapeamentos entre conceitos - ponte semântica (ConceptBridge).....	86
FIG. 5.10	Exemplo de mapeamentos entre conceitos - ponte semântica (ConceptBridge).....	87
FIG. 5.11	Exemplo de mapeamentos entre conceitos - ponte semântica (ConceptBridge).....	87
FIG. 5.12	Exemplo de mapeamentos entre conceitos - ponte semântica (ConceptBridge).....	87
FIG. 5.13	Representação dos mapeamentos entre conceitos através de um grafo.....	90
FIG. 5.14	Extrato da ontologia CPAD.....	94
FIG. 5.15	Extrato da ontologia LEGISMAR.....	94
FIG. 5.16	Representação final da OE.....	95
FIG. 5.17	Representação final da OE em RDF.....	95
FIG. 5.18	Arquitetura lógica da plataforma COPPEER (MIRANDA, 2006).....	97
FIG. 5.19	Janela principal do COPPEER.....	98
FIG. 6.1	Mapa conceitual do sistema CPAD em UML.....	104
FIG. 6.2	Representação gráfica em RDF da Ontologia CPD.....	104
FIG. 6.3	Mapa conceitual do sistema LEGISMAR em UML.....	105
FIG. 6.4	Representação em RDF da Ontologia LEGISMAR.....	105
FIG. 6.5	Mapa conceitual do sistema SIGDEM em UML.....	106
FIG. 6.6	Representação em RDF da Ontologia SIGDEM.....	107
FIG. 6.7	Mapa conceitual do sistema PESSOAL em UML.....	107
FIG. 6.8	Representação em RDF da Ontologia PESSOAL.....	108
FIG. 6.9	Interface da ferramenta MAFRA para definição nome do arquivo de mapeamento.....	109
FIG. 6.10	Representação das associações criadas pela interface do MAFRA.....	110
FIG. 6.11	Representação das pontes geradas pela interface do MAFRA.....	110
FIG. 6.12	Mapeamento entre a ontologia CPAD e a ontologia LEGISMAR.....	111

FIG. 6.13	Mapeamento entre a ontologia CPAD e a ontologia PESSOAL.....	112
FIG. 6.14	Mapeamentos representados em SBO/RDF.....	113
FIG. 6.15	Mapeamentos representados em SBO/RDF.....	113
FIG. 6.16	Informação de mapeamento entre as ontologias CPAD PESSOAL.....	114
FIG. 6.17	Grafo construído com a informação de mapeamento entre as ontologias CPAD e PESSOAL.....	114
FIG. 6.18	Informação de mapeamento entre as ontologias LEGISMAR e CPAD.....	114
FIG. 6.19	Grafo construído com a inclusão da informação de mapeamento entre as ontologias LEGISMAR e CPAD.....	115
FIG. 6.20	Informação de mapeamento entre as ontologias PESSOAL e CPAD.....	115
FIG. 6.21	Grafo construído com a inclusão da informação de mapeamento entre as ontologias CPAD e PESSOAL.....	115
FIG. 6.22	Grafo com as informações dos mapeamentos entre ontologias.....	115
FIG. 6.23	Informação de mapeamento entre as ontologias SIGDEM e LEGISMAR.....	116
FIG. 6.24	Grafo construído com a inclusão da informação de mapeamento entre as ontologias SIGDEM e LEGISMAR.....	116
FIG. 6.25	Exemplo de um mapeamento entre conceitos - ponte semântica.....	116
FIG. 6.26	Grafo construído com a informação dos mapeamento entre os conceitos.....	116
FIG. 6.27	Exemplo de um mapeamento entre conceitos - ponte semântica.....	117
FIG. 6.28	Grafo construído com a inclusão da informação de mapeamento entre os.....	117
FIG. 6.29	Exemplo de um mapeamento entre conceitos - ponte semântica.....	117
FIG. 6.30	Grafo construído com a inclusão da informação de mapeamento entre os conceitos.....	117
FIG. 6.31	Exemplo de um mapeamento entre conceitos - ponte semântica.....	117
FIG. 6.32	Grafo construído com a inclusão da informação de mapeamento entre os conceitos.....	118
FIG. 6.33	Exemplo de um mapeamento entre conceitos - ponte semântica.....	118
FIG. 6.34	Grafo construído com a inclusão da informação de mapeamento entre os conceitos.....	118
FIG. 6.35	Exemplo de um mapeamento entre conceitos - ponte semântica.....	118
FIG. 6.36	Grafo construído com a inclusão da informação de mapeamento entre conceitos.....	118
FIG. 6.37	Exemplo de um mapeamento entre conceitos - ponte semântica.....	119

FIG. 6.38	Grafo construído com a inclusão da informação de mapeamento entre conceitos.....	119
FIG. 6.39	Exemplo de um mapeamento entre conceitos - ponte semântica.....	119
FIG. 6.40	Grafo construído com a inclusão da informação de mapeamento entre conceitos.....	119
FIG. 6.41	Subgrafo com as informações de <i>ranking</i>	120
FIG. 6.42	Subgrafo com as informações de <i>ranking</i> dos conceitos.....	120
FIG. 6.43	Representação final da OE.....	121
FIG. 6.44	Representação final da OE em RDF.....	121
FIG. 6.45	Arquitetura da Bancada de teste.....	129
FIG. 6.46	Representação XML da combinação entre ontologias.....	130
FIG. 6.47	Resultados do experimento 1.....	133
FIG. 6.48	Resultados do experimento 2.....	134
FIG. 6.49	Resultados do experimento 3.....	134

LISTA DE TABELAS

TAB. 5.1	Matriz de adjacência representativa dos mapeamentos existentes em um ponto.....	83
TAB. 5.2	Matriz de adjacência representativa dos mapeamentos entre conceitos existentes em um ponto.....	89
TAB 5.3	Matriz de adjacência representativa dos mapeamentos entre conceitos existentes em um ponto, considerando a direção dos mapeamentos.....	92

RESUMO

O aumento exponencial de informações vem exigindo das organizações de grande porte soluções inovadoras para a integração destas informações. Muitas destas abordagens envolvem a utilização de ontologias, com o objetivo de permitir uma visão global da organização. Porém a construção de uma ontologia é uma tarefa complexa e custosa.

As organizações de grande porte geralmente configuram-se como ambientes autônomos, onde há um alto grau de heterogeneidade, e as informações são representadas das mais diversas formas. Arquiteturas ponto-a-ponto (P2P) oferecem uma solução interessante para estes ambientes na medida em que mantêm a autonomia, mas facilitam a integração entre os diversos pontos da organização. As abordagens P2P apontam para o uso de mapeamentos entre ontologias, entretanto, nenhuma das abordagens propostas oferece a possibilidade de geração de uma ontologia que represente a organização como um todo, ou seja, uma visão global da mesma.

Inspirado nos chamados sistemas emergentes, este trabalho propõe uma abordagem denominada Ontologias Emergentes (OE) para a geração de uma ontologia da organização. Nesta abordagem, considera-se que a informação destas organizações estejam distribuídas em vários pontos, e que em cada ponto esta informação é representada através de ontologias. Além disso, à medida que estes pontos precisam trocar informações, mapeamentos P2P vão surgindo entre as ontologias. Neste contexto, foi definido um conjunto de heurísticas que são aplicadas ao conjunto de mapeamentos P2P para a geração da OE, representando a visão global da organização.

Assim, dentre as principais contribuições deste trabalho destacam-se: a proposta de extensão de uma arquitetura P2P, acrescentando-se a funcionalidade necessária para a geração da Ontologia Emergente; a definição e formalização das heurísticas, viabilizando e facilitando, conseqüentemente, a implementação de um protótipo; e o protótipo OntoEmerge propriamente dito.

Para avaliar a usabilidade do protótipo foram apresentados um exemplo de uso e realizados alguns testes. No exemplo de uso foram utilizadas 4 ontologias de aplicação, baseadas em sistemas de gerência de documentos. Para os testes foram utilizadas 31 ontologias da área bibliográfica, obtendo resultados favoráveis à adoção da abordagem de Ontologias Emergentes.

ABSTRACT

The exponential growth of information comes from large-scale organizations requiring innovative solutions for the integration of this information. Many of these approaches involve the use of ontologies, in order to allow an overview of the organization. But the construction of an ontology is difficult and expensive.

Generally large-scale organisations constitute themselves as autonomous environments where there is a high degree of heterogeneity, and the information is represented in the most diverse ways. Peer-to-peer architectures (P2P) offer an interesting solution for these environments to the extent that they maintain the autonomy, and facilitate the integration between the various peers of the organization. The P2P approaches usually move towards the use of mappings between ontologies, however, none of these approaches offer the possibility of generating an ontology representing the organization as a whole, or an integrated view of it. Inspired in the so-called emerging systems, this work suggests an approach called Emerging Ontologies (EO) for the generation of an ontology of the organization. In this approach, it is considered that the information of these organizations are distributed at various peers, and that at each peer, the information is represented by ontologies. Moreover, as these peers need to exchange information, P2P mappings between ontologies arise. In this context, it was defined a set of heuristics that are applied to the set of P2P mappings for the generation of the EO, representing the global vision of the organization.

Thus, among the major contributions of this work: the proposed extension of a P2P architecture, adding to it the necessary functionality for the generation of an Emerging Ontology; the definition and formalization of the heuristics that supports this new functionality, enabling and facilitating, therefore, the implementation of prototype; and the prototype OntoEmerge itself.

To evaluate the usability of the prototype it was presented an example and some tests' results. In the usage example, it was used 4 application ontologies, based on real systems for the management of documents. For the tests, it was used 31 ontologies from of bibliographic area, obtaining results favourable to the adoption of the Emerging Ontology approach.

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

Normalmente as organizações lidam com as mais variadas formas textuais de informação, a exemplo de normas, livros, boletins técnicos, manuais e outros documentos, disponibilizados e distribuídos de forma eletrônica. Estas fontes de informação são classificadas no contexto deste trabalho como *formais*, pois existem regras para sua confecção, armazenamento e acesso. Por outro lado, existem também documentos em estações de trabalho locais, classificados aqui como *informais*, do cotidiano dos usuários. Estes documentos eletrônicos geralmente ficam restritos a um pequeno grupo de pessoas, e a informação valiosa aí contida muitas vezes acaba se perdendo, ora por conta de mudanças estruturais da organização (transferência de pessoal entre organizações ou setores), ora por questões pessoais (demissões, aposentadorias, etc.). Neste contexto, o armazenamento, compartilhamento e troca de informações *formais* e *informais* não é possível, devido a não haver comprometimento do pessoal da organização em torná-los explícitos de maneira que possam ser acessíveis.

Em muitas organizações, por exemplo, existe constantemente a saída e entrada de pessoal por diversos motivos, afastando-se de suas funções habituais ou incluindo em outras funções que não tinham relacionamento com a anterior. Quando há tempo hábil, o servidor que deixa um determinado tipo de trabalho transmite algumas normas, experiências e conhecimentos adquiridos no dia-a-dia, mas às vezes este tipo de situação não ocorre, levando o recém chegado, por si só, verificar quais as regras e procedimentos necessários devem ser realizados. Por exemplo, para uma determinada pessoa que tem o mínimo de experiência na área de finanças, será muito fácil saber qual a fonte de consulta necessária para auxiliar na resolução de determinado problema. Porém, para quem desconhece ou tem pouco conhecimento, as dificuldades são grandes e o tempo para resolução de um problema pode levar dias, o que não é bem visto em uma organização militar (OM). Geralmente, para resolver um problema que não seja parte integrante da área ou do domínio de conhecimento de um determinado servidor

público, é necessária a busca de informações através de *e-mails* e contatos telefônicos a determinadas pessoas dominadores do assunto, quando estão disponíveis ou mesmo existirem. Em resumo, a solução de um problema ou a realização de uma tarefa, sempre requerem um conhecimento prévio do domínio em questão, já que não existem ferramentas capazes de integrar todo o conhecimento de uma organização, quer inferindo ou sugerindo soluções possíveis para o problema.

Assim, as tarefas de criar e implantar processos capazes de gerar, armazenar, organizar, disseminar e aplicar o conhecimento produzido e utilizado na OM de modo sistemático, explícito, confiável e acessível à comunidade da organização tornam-se um desafio.

As soluções que vêm sendo adotadas para a representação da informação e do conhecimento das organizações costumam ser baseadas no uso de ontologias. Segundo (GRUBER, 1995), uma ontologia é uma especificação explícita de uma conceitualização compartilhada. O uso de ontologias permite representações mais semânticas das informações e, conseqüentemente, facilita o compartilhamento destas informações. No entanto, a construção de uma ontologia a nível organizacional é custosa e complicada. Para construí-la, é necessário contar com equipes de consultores especializados, mas que costumam adotar uma abordagem centralizada, com pouco suporte automatizado. Além disso, a característica dinâmica das organizações pode levar a uma natural desatualização da ontologia, assim que uma primeira versão da mesma é disponibilizada para uso. Desta forma, faz-se necessário desenvolver mecanismos que auxiliem a concepção de uma ontologia organizacional de forma mais ágil e descentralizada, de modo a refletir o dinamismo da representação informação.

Algumas arquiteturas distribuídas investigadas na literatura, voltadas para a disseminação de informações na Web, e particularmente, adequadas ao uso das tecnologias mais recentes da Web Semântica, destacam duas categorias: Grade e Ponto-a-Ponto (P2P). A arquitetura de grade é utilizada em aplicações de alto desempenho e com poucas organizações e servidores. Já as arquiteturas P2P são caracterizadas pelo compartilhamento de recursos computacionais e serviços através da comunicação direta e descentralizada entre os sistemas envolvidos, tendo a participação de uma quantidade maior de usuários (OOI et al, 2003). Os recursos e serviços destas arquiteturas incluem, entre outras coisas, a troca de informações. Algumas abordagens tais como (PALMA et al., 2005) e (BROEKSTRA et al. 2004) utilizam ontologias para facilitar esta troca de informações. Entretanto, embora estas abordagens ofereçam um modo descentralizado de representação da informação, costumam considerar ontologias similares,

ou até mesmo idênticas, residentes em cada ponto. Além disso, não possuem como objetivo a geração de uma ontologia representativa de todos os pontos.

1.2 OBJETIVO

Com o objetivo de apoiar a construção mais ágil e descentralizada de uma visão integrada da organização, este trabalho propõe a extensão de uma arquitetura P2P baseada em ontologias. Sabendo-se que uma organização é usualmente composta por equipes de trabalho (projetos, comitês, etc.), envolvendo grupos de pessoas, considera-se que cada ponto poderia representar um destes grupos dentro da organização, e ainda, que cada grupo teria a representação de suas informações ou conhecimento através de uma ontologia local. Considerando que as equipes interagem e precisam trocar informações, é natural que haja uma necessidade de mapeamento entre as ontologias de cada ponto correspondente. Mecanismos para interoperabilidade de ontologias (FELICÍSSIMO, 2004), (FALCONER et al, 2006), (NOY e MUSEN, 2003), (OMELAYENKO, 2002) e (NUNO SILVA, 2004) podem prover estes mapeamentos.

Sendo assim, estas informações valiosas sobre os conceitos compartilhados a partir destes mapeamentos, facilitariam o surgimento de uma visão integrada da organização, aqui denominada de Ontologia Emergente (OE). Conseqüentemente, uma visão local e global das informações da organização se tornaria disponível para buscas e trocas de informações, evitando a perda das mesmas.

Visando alcançar este objetivo, uma arquitetura e um protótipo foram desenvolvidos, levando em consideração tecnologias já existentes, como dos sistemas P2P e da Web Semântica.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação encontra-se organizada em 7 capítulos, a saber.

O capítulo 1 apresenta a introdução, expondo a motivação, objetivo e organização dessa dissertação.

O capítulo 2 faz um levantamento das tecnologias relacionadas à interoperabilidade de ontologias, com destaque para as principais abordagens e ferramentas relacionadas.

O capítulo 3 faz um estudo sobre as tecnologias de distribuição e integração de dados, abordando alguns aspectos relacionados à Web Semântica.

O capítulo 4 apresenta a abordagem de Ontologias Emergentes (OE), destacando a arquitetura proposta, com a descrição de seus módulos e heurísticas adotadas para a geração da OE.

O capítulo 5 descreve os detalhes da implementação do protótipo *OntoEmerge*.

O capítulo 6 descreve o exemplo de uso do protótipo, desde os mapeamentos até a geração da OE, tendo como base quatro ontologias. Para validar a OE foram realizados alguns experimentos que são descritos em detalhe, juntamente com os resultados obtidos.

Finalmente, no capítulo 7, são apresentadas as conclusões dessa dissertação, suas contribuições e algumas sugestões para trabalhos futuros.

2 INTEROPERABILIDADE DE ONTOLOGIAS

Ontologias suportam o compartilhamento e reuso do conhecimento. No entanto, para que trabalhem em conjunto trocando e compartilhando automaticamente as informações representadas, é necessário que possam interoperar. A literatura aponta um conjunto expressivo de ferramentas que suportam a interoperabilidade de ontologias (MAEDCHE et al, 2002), (FALCONER et al, 2006) e (FARQUHAR et al., 1996). Num documento recente, diversas abordagens que lidam com interoperabilidade entre ontologias foram investigadas (LACHTIM et al. ,2008). Este relatório descreve exhaustivamente uma enorme gama de ferramentas, razão pela qual neste capítulo serão tecidas somente breves considerações sobre o tema, e apresentadas algumas tecnologias que tratam de interoperabilidade de ontologias, com ênfase para suas características relevantes, tendo como base o documento supracitado.

2.1 ONTOLOGIAS

Com o advento da Web Semântica (WS) (BERNERS-LEE et al., 2001) a W3C (World Wide Web Consortium) vem trabalhando arduamente no desenvolvimento de tecnologias avançadas, que visam à representação estrutural e semântica dos recursos na Web. Essas tecnologias, aliadas à teoria de domínios ou ontologias, permitem oferecer um serviço Web com um nível maior de qualidade.

Algumas tecnologias foram desenvolvidas, tais como: linguagens de transporte de dados (ex. XML), linguagem para representação da estrutura desses dados (ex. RDF) e linguagens para representação da semântica desses dados (ex. OWL), explicitando restrições sobre a semântica do mundo real. Vinculado a essa gama de linguagens, há também as ferramentas para manipulação das ontologias, a exemplo do OntoEdit (SURE et al., 2002), do KAON (KAON) e do Protégé (PROTÉGÉ), entre outras.

As ontologias são de grande importância na WS, pois conseguem embutir significado sem ambigüidade através de vocabulários, permitindo interconexões semânticas entre os termos e regras de inferência e lógica sobre um determinado domínio de conhecimento. As ontologias

facilitam a interpretação e recuperação das informações, ao mesmo tempo em que viabilizam o intercâmbio entre sistemas. Possibilitam também um mecanismo de pesquisa mais apurado e restrito às informações realmente relevantes, automação de tarefas que exijam raciocínio e sugestão de opções e caminhos, auxiliando o usuário no alcance dos seus objetivos. No processo de integração de dados, elas possuem um papel fundamental dentre os quais se pode destacar:

- Explicitação do conteúdo das fontes de dados. Consiste em descrever a semântica referente aos dados das diversas fontes de dados, garantindo assim, um vocabulário compreensível a todos;
- Construção de um modelo de consulta global. Consiste em criar uma grande estrutura que pode ser utilizada para descobrir o que realmente os objetos significam. Auxilia na reescrita da consulta para as fontes ou pontos relevantes;
- Auxílio lógico. Consiste em definir regras de mapeamentos e restrições de integridade que auxiliem o sistema de integração de dados na extração dos dados e no processamento de consultas, através da derivação lógica destas regras.

Como se pode ver, o uso de ontologias no processo de integração de dados é fundamental para a compreensão semântica dos dados oriundos das diversas fontes. Além de possibilitar e facilitar o processo de integração, este representa uma garantia do aumento da qualidade e precisão na resolução de conflitos e das consultas e esquemas gerados.

Uma das definições de ontologia freqüentemente encontrada e estudada na literatura é: ontologia é uma "especificação formal (BORST, 1997) e explícita (GRUBER, 1993) de uma conceitualização compartilhada".

- *Formal*: refere-se ao fato de que uma ontologia deveria ser "lida" por máquina;
- *Compartilhada*: ontologia deve refletir o conhecimento consensual, aceito por um grupo;
- *Explícita*: significa que o tipo de conceitos usados e restrições sobre seu uso são explicitamente definidos; e
- *Conceitualização*: é basicamente uma idéia do mundo que uma pessoa ou um grupo de pessoas podem ter.

Existem várias classificações diferentes de ontologias fornecidas por diferentes autores. Em (GUARINO, 1998) é proposta a classificação que usa a generalidade como principal critério para o desenvolvimento dos diferentes tipos de ontologias:

- **ontologias de alto nível (topo ou gerais):** descrevem conceitos de forma bem geral para compreensão dos aspectos do mundo, como espaço, tempo, material, objeto, evento, ação, etc., os quais são independentes de um problema ou domínio particular. Podem ser compartilhadas por grandes comunidades de usuários;
- **ontologias de domínio:** descrevem o vocabulário relacionado a um domínio genérico (como medicina, automóveis, ou ainda uma organização como por exemplo a Marinha), através da especialização da ontologia de topo;
- **ontologias de tarefas:** descrevem o vocabulário relacionado a uma tarefa genérica (como diagnóstico ou vendas), através da especialização da ontologia de topo; e
- **ontologias de aplicação:** descrevem conceitos dependendo do domínio e de tarefas particulares. Estes conceitos, freqüentemente, correspondem a papéis desempenhados por entidades do domínio, quando na realização de certas tarefas. Por exemplo, ajuda para o diagnóstico de doenças mentais.

A FIG. 2.1 mostra a relação entre estas ontologias. Os conceitos de uma ontologia de domínio ou de tarefa são especializações dos conceitos das ontologias de topo. Assim, as ontologias de topo possuem maior capacidade de reuso, enquanto que as ontologias de aplicação são as que possuem menor capacidade de reuso, por definir conceitos relativos a uma aplicação específica.

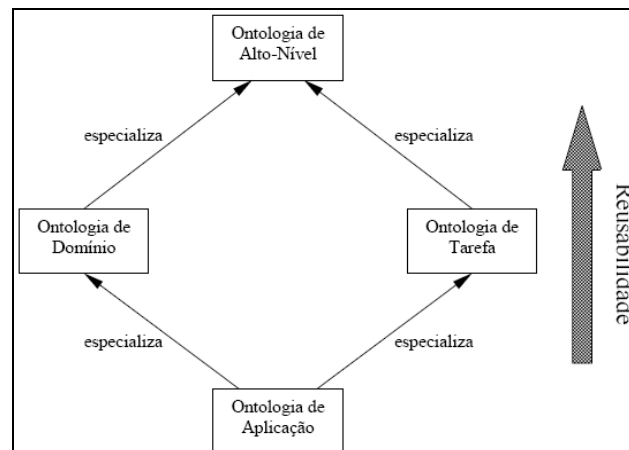


FIG. 2.1 Classificação das ontologias (GUARINO, 1998)

2.2 ABORDAGENS PARA INTEROPERABILIDADE DE ONTOLOGIAS

Para que as informações disponíveis na *Web* sejam utilizadas pelos diferentes sistemas, é necessário, em um primeiro momento, que estas sejam localizadas, acessadas, compreendidas e processadas pelos mesmos. Em um segundo momento, devido à heterogeneidade do conteúdo dessas informações, que podem ser classificadas segundo os aspectos estrutural, sintático e semântico (SHETH, 1999), é preciso compatibilizar esses conteúdos entre si.

Heterogeneidade estrutural refere-se à capacidade dos sistemas representarem suas entradas segundo diversas representações e modelos. Existem mecanismos que avaliam a heterogeneidade estrutural segundo as hierarquias de conceitos e relacionamentos encontrados nas ontologias, visando encontrar estruturas equivalentes. No caso mais simples, uma ontologia descreve uma hierarquia de conceitos relacionados por meio de regras, enquanto que, em casos mais sofisticados, são adicionados axiomas adequados para expressar relacionamentos mais complexos e para restringir a interpretação desejada de seus conceitos (GUARINO, 1998).

A *heterogeneidade sintática* tem como elementos principais: metadados, termos e linguagens que são utilizados na representação do conhecimento. A heterogeneidade pode ser resolvida com a utilização, por exemplo, de dicionários de termos léxicos.

Entende-se por interoperabilidade semântica a capacidade de dois ou mais sistemas heterogêneos e distribuídos trabalharem em conjunto, compartilhando as informações entre eles com o entendimento comum de seu significado (BURANARACH, 2004). Segundo (HAKIMPOUR & GEPPERT, 2004), semântica é a interpretação atribuída ao dado por diversas pessoas, sendo que diferentes interpretações de dados causam a *heterogeneidade semântica*.

Dessa forma, tentando ultrapassar alguns aspectos dos conflitos mencionados acima, sistemas de cálculo de similaridade surgem como um passo prévio de formalização das relações entre ontologias. Assim, para que informações oriundas de ontologias diferentes sejam utilizadas, é necessário prever mecanismos eficientes, capazes de suportar os diversos níveis de interoperabilidade (FELICISSIMO, 2004). Para que isto ocorra, devem ser aplicados mecanismos para efetuar o cálculo de similaridades entre entidades de ontologias fonte e alvo. A determinação das equivalências entre as ontologias pode ser encontrada com o apoio de algoritmos de casamento, cujo objetivo é encontrar correspondências entre entidades

de ontologias relacionadas. Estas correspondências podem ser usadas por várias tarefas como: navegação na *Web*, tradução de dados, consultas/respostas à base de dados, etc.

Após a determinação dos valores de similaridades entre as entidades das ontologias fonte e alvo, pode-se então aplicar uma das diversas abordagens para realizar a interoperabilidade de ontologias. Na literatura (MAEDCHE et al, 2002), (FELICISSIMO, 2004), (DOU et al, 2002) e (FARQUHAR et al, 1996), são apontadas quatro abordagens distintas com essa função: combinação (*merge* ou fusão), alinhamento, mapeamento e integração.

- *Combinação de ontologias* - na combinação de ontologias tem-se como resultado a versão das ontologias originais combinadas em uma ontologia única, com todos seus termos juntos e sem a definição clara de suas origens. Normalmente as ontologias originais descrevem domínios similares ou de sobreposição.
- *Alinhamento de ontologias* - no alinhamento de ontologias tem-se como resultado as duas ontologias originais separadas, mas a estas são adicionadas ligações entre seus termos equivalentes. Estas ligações permitem que as ontologias alinhadas reusen informações umas das outras. O alinhamento normalmente é realizado quando as ontologias são de domínios complementares.
- *Mapeamento de ontologias* - no mapeamento de ontologias tem-se como resultado uma estrutura formal com expressões que ligam os termos de uma ontologia aos termos de outra ontologia. Este mapeamento é usado para transferir instâncias de dados, esquemas de integração e de combinação e outras tarefas similares.
- *Integração de ontologias* – três diferentes significados para o termo integração foram identificados em (PINTO et al, 1999):
 - i.* na construção de uma nova ontologia reusando (pela união, extensão, especialização ou adaptação) termos ou conceitos de outras ontologias já existentes;
 - ii.* na construção de ontologias utilizando o processo de "merge" de várias ontologias, tendo como resultado final uma única ontologia que representa todas que sofreram o "merge". Esse tipo de integração equivale à abordagem já descrita, de combinação de ontologias; e
 - iii.* na construção de uma aplicação que usa uma ou mais ontologias.

No escopo deste trabalho chamaremos de integração de ontologia a identificada em *i*, onde o reuso de conceitos é o principal alvo. Nessa abordagem é possível identificar as regiões que foram criadas a partir das ontologias originais.

2.3 FERRAMENTAS DE CÁLCULO DE SIMILARIDADES ENTRE ONTOLOGIAS

A similaridade é essencial em abordagens que visam o estabelecimento de uma relação formal entre as entidades da ontologia, tais como alinhamento, mapeamento, combinação e integração. O casamento de entidade possui um papel decisivo e atua como um passo anterior à aplicação de uma das abordagens descritas na seção anterior. Assim, quanto mais aprimorado for este mecanismo, melhor serão os resultados fornecidos pelos mecanismos de interoperabilidade, proporcionando aos sistemas que utilizam estas abordagens maior grau de automatização.

Na literatura existe uma variedade de sistemas com o propósito de encontrar similaridades entre entidades de ontologias. Alguns fazem uso de recursos de vocabulários controlados, tais como, CMS (KALFOGLOU & HU, 2005), AUTOMS (KOTIS et al, 2006) e S-MATCH (GIUNCHIGLIA et al, 2004), que utilizam heurísticas relacionadas com seus conceitos e propriedades, enquanto outros utilizam técnicas que envolvem a estrutura das ontologias, a exemplo do GLUE (DOAN et al, 2004), OMEN (MITRA et al, 2004), CMS (KALFOGLOU & HU, 2005), AUTOMS (KOTIS et al, 2006) e S-MATCH (GIUNCHIGLIA et al, 2004), sobrepondo conflitos como os mencionados em (XIAOMENG, 2004). Apesar dessas técnicas, a realidade é que estas ferramentas ainda não conseguem resolver todos os conflitos existentes em decorrência do processo de casamento.

2.4 TECNOLOGIAS PARA INTEROPERABILIDADE ENTRE ONTOLOGIAS

A necessidade de reusar e/ou compartilhar informação é facilmente compreendida, porém transformar esse conceito em fato concreto exige um arcabouço de padrões pré-definidos tais como métodos, ferramentas e conhecimento teórico sobre a manipulação das ontologias. Sem a disponibilidade deste ferramental, a interoperabilidade de sistemas baseados em ontologias ficaria comprometida. Existem diversas ferramentas, métodos e sistemas disponíveis para apoiar as diversas abordagens de interoperabilidade entre ontologias.

No relatório técnico (LACHTIM et al. ,2008) são apresentadas algumas ferramentas para lidar com a interoperabilidade de ontologias, destacando suas características e

funcionalidades. Nele é feito um estudo comparativo entre as ferramentas, que são classificadas da seguinte forma:

- *Mapeamento* - MAFRA (MAEDCHE et al, 2002), PROMPT (FALCONER et al, 2006) e RDFT (OMELAYENKO, 2002);
- *Alinhamento* - PROMPT (NOY & MUSEN, 2000) e CATO (FELICISSIMO, 2004);
- *Combinação* - FCA-Merge (MAEDCHE & STUMME, 2001) e ONTOMERGE (DOU et al, 2002); e
- *Integração* - ONTOLINGUA (FARQUHAR et al, 1996).

Nas próximas seções será apresentada, sucintamente, uma ferramenta de cada abordagem. As demais estão descritas com maiores detalhes no relatório citado (LACHTIM et al., 2008).

2.4.1 TECNOLOGIAS PARA MAPEAMENTO

Três ferramentas relacionadas a mapeamentos de ontologias foram investigadas: MAFRA (MAEDCHE et al, 2002), PROMPT (FALCONER et al, 2006) e RDFT (OMELAYENKO, 2002), sendo que a ferramenta PROMPT permite também o alinhamento de ontologias. Todas possuem mecanismos que proporcionam uma forma de explicitação dos mapeamentos existentes, representados através de uma linguagem de ontologias (OWL / RDFS).

Devido a importância da ferramenta MAFRA no contexto dessa dissertação, sua descrição mereceu maior destaque nesta seção.

2.4.1.1 MAFRA (MAEDCHE et al, 2002) e (NUNO SILVA, 2004)

MAFRA (*MApping FRamework for distributed ontologies*) é um framework definido para o mapeamento de ontologias distribuídas na Web Semântica. Tem como objetivo detectar automaticamente similaridades entre entidades de duas ontologias distintas, realizando o mapeamento entre elas e transformando instâncias da ontologia fonte em instâncias da ontologia alvo. As ontologias devem ser normalizadas em uma representação uniforme,

eliminando diferenças de sintaxe e tornando as diferenças semânticas mais aparentes. Na arquitetura do MAFRA são identificados dois novos conceitos:

- *Pontes Semânticas*: trata-se de uma abordagem baseada em *Serviços de Transformação*, na qual uma ponte semântica é uma representação declarativa de uma relação semântica entre entidades de duas ontologias. Esta ponte fornece os mecanismos necessários para transformar instâncias e propriedades de uma ontologia fonte em instâncias e propriedades de uma ontologia alvo. Para isto, cada ponte semântica possui *Serviços de Transformação* associados.
- *Semantic Bridging Ontology (SBO)*: toda informação relativa ao processo de mapeamento é instanciada numa *SBO* (FIG. 2.2), que é uma ontologia para mapeamento de ontologias, ou seja, é uma meta-ontologia que especifica, classifica e descreve os tipos de relação de mapeamento de ontologia.

O MAFRA está em desenvolvimento como um módulo adicional à arquitetura KAON (KAON).

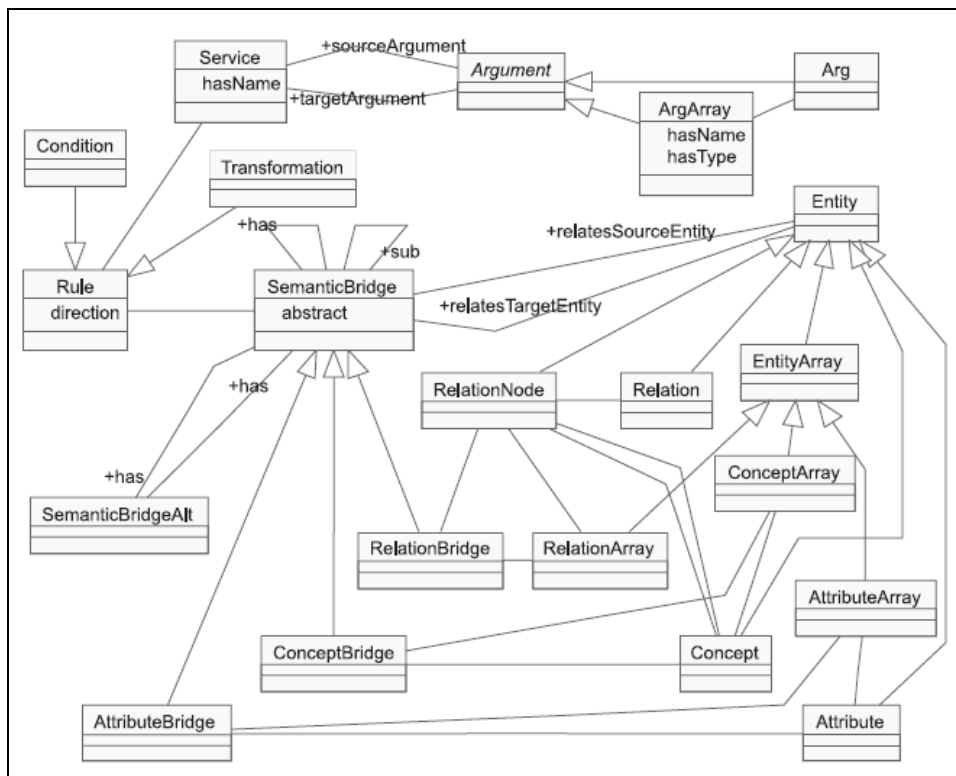


FIG. 2.2 Estrutura da SBO (MAEDCHE et al, 2002)

O framework do MAFRA consiste de cinco módulos horizontais e quatro componentes verticais. Estes são executados ao longo do processo de mapeamento, e interagem com os módulos horizontais, a saber:

- **Lift e normalização:** este módulo é utilizado para transformar os dados em um mesmo nível de representação sintática e estrutural da linguagem. As ontologias devem ser normalizadas para uma representação uniforme em RDF, eliminando as diferenças de sintaxe e tornando as diferenças semânticas entre ontologias fonte e alvo mais aparentes;
- **Similaridade:** este módulo estabelece similaridades entre entidades de ontologias fonte e alvo;
- **Pontes Semânticas:** baseada nas similaridades computadas previamente, este é responsável por estabelecer correspondências entre entidades da ontologia fonte e alvo. O papel da ponte semântica é encapsular toda a informação necessária para transformar instâncias de uma ontologia fonte em instâncias de uma ontologia alvo;
- **Execução:** este módulo transforma instâncias de uma ontologia fonte em instâncias de uma ontologia alvo, de acordo com as pontes semânticas;
- **Pós-Processamento:** este módulo recebe como entrada os resultados do módulo de execução, melhorando a qualidade dos resultados transformados.

A dimensão vertical do MAFRA contém módulos que interagem com os módulos horizontais durante todo o processo de mapeamento. Existem módulos que possuem estratégias de sincronização de mudanças para permitir a evolução de ontologias. Uns estabelecem um consenso de construção das pontes semânticas entre duas comunidades que participam do processo de mapeamento, enquanto outros são responsáveis pela utilização de um conhecimento adicional, através do uso de glossários, vocabulários controlados ou tesouros, para ajudar a identificar sinônimos, tais como o WordNet (WORDNET).

A FIG. 2.3 apresenta uma visão geral da arquitetura conceitual do MAFRA, mostrando a interação entre os diversos módulos.

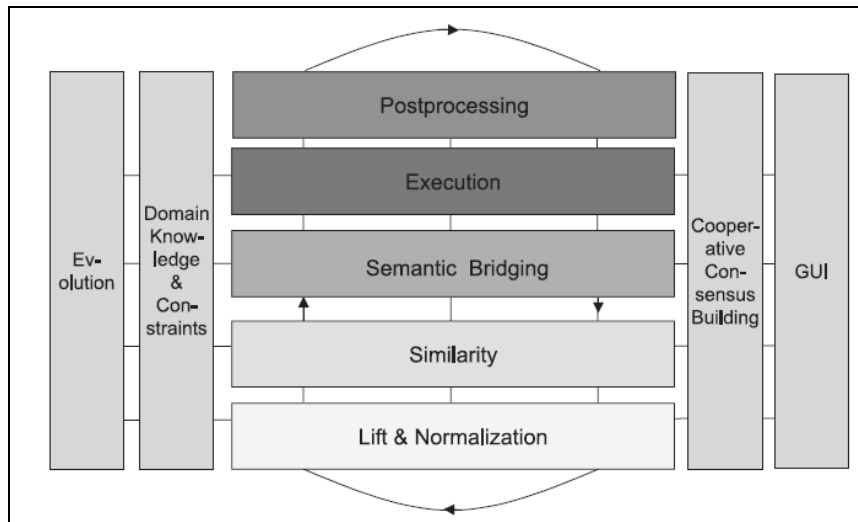


FIG. 2.3 Arquitetura conceitual do MAFRA (MAEDCHE et al, 2002)

2.4.2 TECNOLOGIAS PARA ALINHAMENTO

Para o alinhamento foram estudadas duas ferramentas: PROMPT (NOY & MUSEN, 2000) e CATO (FELICISSIMO, 2004), das quais apenas a segunda será apresentada.

2.4.2.1 CATO (FELICISSIMO, 2004)

CATO (**C**omponente para **A**linhamento **T**axonômico de **O**ntologias) é uma ferramenta desenvolvida na Pontifícia Universidade Católica do Rio de Janeiro, com o objetivo de alinhar automaticamente as taxonomias das ontologias de entrada a serem comparadas. O alinhamento realizado é dividido em três etapas:

- A primeira etapa compara lexicamente os conceitos das ontologias de entrada e usa um mecanismo de poda estrutural dos conceitos associados como condição de parada;
- A segunda etapa compara estruturalmente as hierarquias das ontologias, identificando as similaridades entre suas sub-árvores comuns;
- A terceira etapa refina os resultados da etapa anterior, classificando os conceitos identificados como: similares, bem similares ou pouco similares, de acordo com um percentual de similaridade pré-definido.

A FIG. 2.4 apresenta uma visão geral da estratégia proposta pelo CATO.

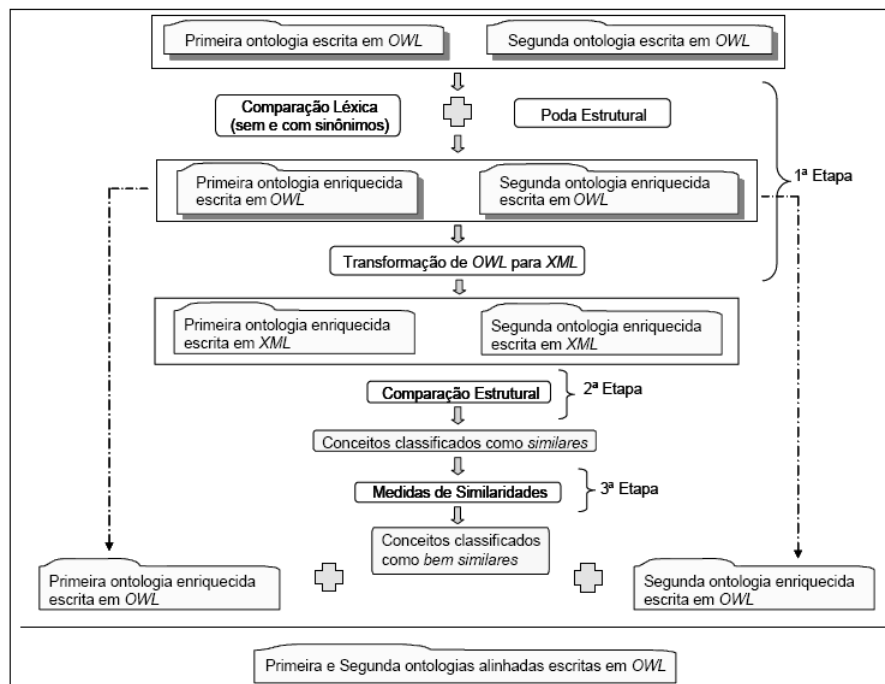


FIG. 2.4 Estratégia para alinhamento taxonômico de ontologias (FELICISSIMO, 2004)

No final da execução das três etapas da estratégia, as informações de equivalência de conceitos bem similares são adicionadas às ontologias resultantes da primeira etapa. Após esta adição, as ontologias alinhadas são unidas em uma ontologia única. Esta ontologia única é o resultado final da estratégia. Vale observar que, conforme descrito na seção 2.2, a existência de uma única ontologia no final de um processo é característica da integração e/ou combinação de ontologias. Porém, a justificativa dos autores em dizer que, apesar desta junção das ontologias, o processo é declarado como alinhamento, é o fato das ontologias alinhadas pela estratégia continuarem a serem reconhecidas pela identificação de seus *namespaces*. Existe a ligação entre os conceitos equivalentes na ontologia única, permitindo a reutilização e o compartilhamento de suas informações comuns, características do mecanismo de alinhamento de ontologias.

2.4.3 TECNOLOGIAS PARA COMBINAÇÃO OU FUSÃO

As ferramentas relacionadas à combinação/fusão de ontologias são FCA-Merge (MAEDCHE & STUMME, 2001) e ONTOMERGE (DOU et al, 2002), sendo esta última descrita a seguir.

2.4.3.1 ONTOMERGE (DOU et al, 2002)

O ONTOMERGE (Ontology Translation by Merging Ontologies) está disponibilizado *on-line*¹ para a realização de combinação de ontologias. É um serviço semi-automático que dá suporte a humanos e a agentes de software na tarefa de encontrar diferenças de notação entre duas ontologias de domínios sobrepostos. É constituído das seguintes etapas:

- O primeiro passo realiza uma tradução das ontologias a serem combinadas em uma linguagem de representação própria, a Web-PDDL, que é fortemente tipada segundo uma linguagem de primeira ordem para aplicação Web;
- A partir daí um engenheiro de ontologia define os axiomas denominados pontes, que são axiomas que contêm termos de ambas as ontologias. Em seguida ocorre o processo de tradução das instâncias: todas as instâncias das ontologias fontes são consideradas no processo de *merge*;
- A partir das ontologias fontes e das pontes entre os elementos de ambas as ontologias definidas, um motor de inferência, chamado *OntoEngine* realiza o merge das ontologias (DOU et al, 2005).

O OntoMerge possui em sua arquitetura o OntoEngine e o PDDAML. O PDDAML realiza a tradução sintática das entradas e saídas: DAML de/para Web-PDDL, respectivamente. Na FIG. 2.5 é possível observar sua arquitetura.

¹ <http://aimlab.cs.uoregon.edu/ontomerge/ontoMerge.html>

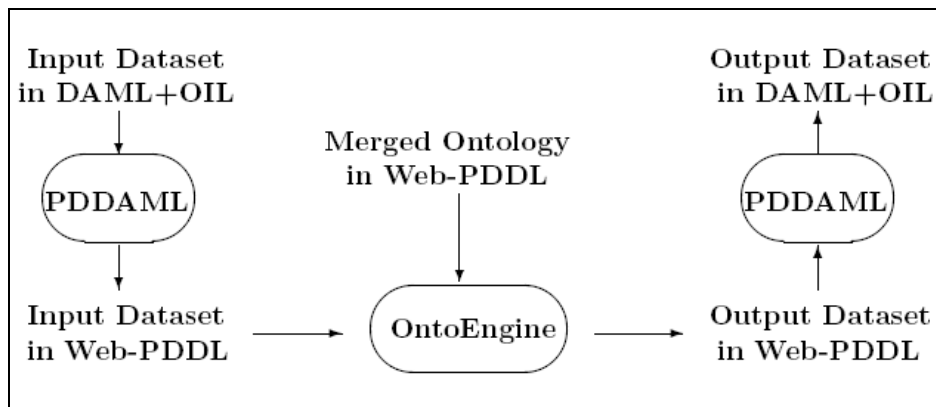


FIG. 2.5 Arquitetura do OntoMerge (DOU et al, 2002)

2.4.4 TECNOLOGIAS PARA INTEGRAÇÃO

Nesta seção, é apresentada uma ferramenta relacionada à integração de ontologias. Pelo fato da integração possuir um processo de unificação por composição, é possível identificar as regiões que foram criadas a partir das ontologias originais.

Em algumas metodologias para construção de ontologias, a integração é reconhecida como parte integrante do processo de desenvolvimento de ontologias (USCHOLD et al. 1995), (GRÜNINGER e FOX 1995) e (FERNANDEZ et al. 1997). Tendo como base a definição *i* adotada por (PINTO et al. 1999) (seção 2.2), foram investigadas diversas ferramentas de integração, porém, somente o ONTOLINGUA se adequou a este contexto.

2.4.4.1 ONTOLINGUA (FARQUHAR et al, 1996)

Ontolingua é um ambiente para desenvolvimento de ontologias. Proporciona uma suíte de ferramentas para criação de ontologias, chamadas de ferramentas de autoria, e uma biblioteca com ontologias modulares e reutilizáveis. As ferramentas de autoria permitem a criação de uma ontologia simplesmente através do "link" de objetos.

As ontologias da biblioteca podem ser combinadas e estendidas para a criação de novas ontologias.

O Ontolingua possui características que suportam a inclusão cíclica de uma ontologia em outra, isto é, uma ontologia A incluída em B, B incluída em C e C incluída em A. Permite aos usuários estenderem ontologias de múltiplas maneiras, tais como: adicionando simples hipóteses, ou restrição de domínio, ou estendendo o domínio de diversos operadores.

O ambiente disponibiliza uma interface que possibilita a qualquer usuário da *Web* acessar as ontologias criadas e disponíveis e criar novas ontologias.

A arquitetura do servidor Ontolingua está representada na FIG. 2.6.

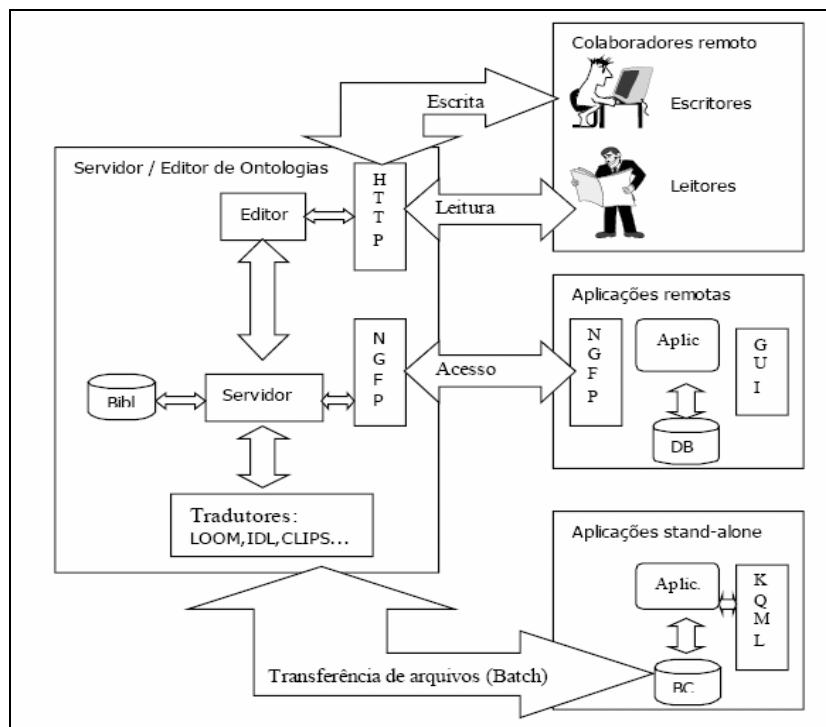


FIG. 2.6 Arquitetura do Ontolingua (FARQUHAR et al, 1996)

Os colaboradores remotos lêem e contribuem com novos conhecimentos para a biblioteca digital de ontologias em texto, mantida em HTML.

As aplicações podem usar as ontologias de duas maneiras: fazendo consultas ou atualizações usando o *NGFP*, uma API (*Application Programming Interface*) que estende o *Protocolo de Frames Genéricos (GFP)* ou mandando e recebendo conhecimento das bases através da linguagem *KIF (Knowledge Interchange Format)*.

Usuários podem traduzir uma ontologia em um formato usado por uma aplicação específica. Por exemplo, uma linguagem de definição de interface (*IDL*) pode produzir um arquivo de cabeçalho que um programa *CORBA* compatível pode usar para interagir com o *object request broker (ORB)*.

2.5 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentadas algumas abordagens de interoperabilidade de ontologias: mapeamento, alinhamento, combinação e integração.

O conceito de integração de ontologias é definido de diferentes formas, mas no contexto aqui proposto foi adotada a definição encontrada em (PINTO et al, 1999), onde a integração é um processo de refinamento, e deve ser construída a partir da reutilização de outras ontologias. A integração não leva em consideração as informações dos mapeamentos realizados entre duas ontologias.

O conceito de mapeamento também é pouco padronizado na literatura. Neste trabalho utilizamos a abordagem utilizada pelo MAFRA, denominada SBO (MAEDCHE et al, 2002).

As ferramentas que dão suporte as abordagens supracitadas, não tratam de várias ontologias ao mesmo tempo, somente de um par de ontologias.

As ferramentas de cálculo de similaridade entre entidades de ontologias foram colocadas como uma fase inicial e necessária para as abordagens supracitadas.

No próximo capítulo serão apresentadas abordagens de integração que envolvem sistemas que proporcionam a interoperabilidade de forma autônoma e descentralizada, a exemplo de Grade e sistemas P2P. Estas arquiteturas se mostram adequadas à veiculação e interoperação de ontologias na Web Semântica.

3 INTEGRAÇÃO DE DADOS EM AMBIENTES AUTÔNOMOS

Com a evolução e a diminuição de custo dos recursos de informática tanto em hardware quanto em software, houve interesse de muitas organizações na descentralização do processamento (no nível do sistema) e na integração dos recursos de informação (no nível lógico) dentro de seus sistemas de banco de dados, aplicações e usuários distribuídos (ELMASRI & NAVATHE, 2005).

O surgimento da Web fez crescer, ainda mais, e drasticamente a busca por modelos e padrões de representação e manipulação de dados visando integrar várias fontes de informação heterogêneas, visto que a Web está se tornando a principal estrutura para publicação e disseminação de informação em diversas organizações, tanto no nível interno quanto no externo (CASTANO et al., 2002).

A tarefa de integração torna-se complexa quando as fontes a serem integradas constituem-se de um grande volume de dados geograficamente distribuídos, e que são, sobretudo de caráter totalmente heterogêneo, no que diz respeito aos seus esquemas de representação. Vários tipos de heterogeneidade podem ser considerados, os quais incluem diferenças de: hardware e sistema operacional, sistemas de gerenciamento de dados, modelos (dados, esquemas e dados semânticos), *middlewares*, interface do usuário e regras de negócio e restrições de integridade (DITTRICH & ZIEGLER, 2003).

Na literatura, existem várias abordagens relacionadas ao compartilhamento e troca de dados e informação. Por exemplo: banco de dados distribuídos, *multidatabases*, sistema de banco de dados federados (FDBMS), sistemas de consultas mediadas, portais, *datawarehouses*, banco de dados universais, Web Semântica, integração realizada por serviços Web, integração ponto-a-ponto (P2P), grades, sistemas de gerenciamento de workflow (WFMS) e outras. Diversos trabalhos na linha de integração de dados ficaram consagrados na literatura (SHETH & LARSON, 1990), (BRIGHT & HURSON, 1992), (DITTRICH & DOMENIG, 1999), com propostas de taxonomias para classificar sistemas de integração de fontes de dados heterogêneas.

Na Web Semântica, que é uma extensão da Web tradicional, propõe-se a compreensão, estruturação e gerenciamento dos conteúdos armazenados na Web, na forma de texto, som, imagem e gráficos atribuindo um significado, de modo que este seja perceptível tanto pelo

humano como pelo computador, permitindo o trabalho em cooperação (BERNERS-LEE et al., 2001). Esta nova Web é capaz de representar integração entre fontes de informação, que a princípio poderiam não estar relacionadas, ou apenas relacionadas entre sistemas internos de uma empresa, por exemplo. No capítulo 2 são apresentadas algumas considerações sobre Web Semântica, englobando ontologias e a interoperabilidade de ontologias.

No contexto dessa dissertação, o foco é na integração de fontes de dados nas organizações de médio e grande porte. Assim, procurou-se focar nas abordagens em que a integração ocorre em ambientes autônomos, frequentemente encontrados nestas organizações. Neste sentido, destacam-se as redes P2P e Grades (ou *Grids*), que são modelos de computação distribuída que possibilitam a colaboração descentralizada integrando computadores na rede. Estas abordagens são detalhadas a seguir.

3.1 GRADES

Grade é uma plataforma de computação geograficamente distribuída que compreende um conjunto de computadores heterogêneos, onde usuários podem acessar através de uma única interface. As Grades, em sua essência, são sistemas P2P. No futuro, muitos aspectos de sua complexa implementação baseadas em serviços hierárquicos, devem ser mudados. As Grades usam aplicações complexas adicionando de 10 a 1000 nós, descentralizando funcionalidades e evitando gargalos (TALIA & TRIUNFIO, 2003).

Grades Computacionais (do inglês *Grid Computing*) surgiram em meados da década de 90, com o objetivo de viabilizar a execução de aplicações paralelas entre diversas máquinas, podendo ser em rede local ou rede de longa distância, formando uma máquina virtual (CIRNE & SANTOS-NETO, 2005). Tal proposta tinha dois grandes motivos: o de fornecer uma plataforma menos custosa para execução de aplicações distribuídas, comparado aos supercomputadores paralelos; e de possibilitar, através da reunião de recursos dispersos, a execução de aplicações paralelas em uma escala simplesmente impossível em um único supercomputador. Estes recursos podem significar informações, poder computacional, unidades de armazenamento de dados, base de dados, software entre outros.

Em (FOSTER, 2001) define-se uma estrutura em níveis que englobam os requisitos básicos para definir uma arquitetura de grade.

Com a evolução da tecnologia de Grades percebeu-se que a composição automática de um conjunto de recursos para servir uma aplicação criava a oportunidade de oferecer serviços sob demanda, e assim, com a idéia da Grade, seria possível prover diversos serviços computacionais (não somente serviços para computação de alto desempenho) sob demanda (CIRNE & SANTOS-NETO, 2005).

Segundo (DE ROURE et al, 2003), a evolução da Grade se deu em três gerações: na primeira os sistemas envolveram soluções proprietárias para compartilhamento de recursos de alto desempenho; na segunda os sistemas introduziram os *middlewares* para resolver questões relativas à escalabilidade e heterogeneidade, focada em computação em alta escala e grandes volumes de dados; e na terceira e atual, os sistemas adotam abordagens orientadas a serviços (*Web Services*) e utilizam-se de metadados, que as caracterizam de forma independente. Estes sistemas requerem informações como funcionalidades, disponibilidades e interfaces dos vários componentes.

Na terceira geração, a tecnologia de Grade evolui, passando a agregar mais semântica, similar à evolução da Web para a Web semântica. A Grade Semântica faz uso de metadados e ontologias como forma de representar, armazenar, acessar, compartilhar e manter a informação, isto é, a informação é compreendida como dados que possuem um significado.

A Grade e a Web Semântica têm diversas características em comum que são talvez, diferenciadas pela ênfase dada: a Grade é focada tradicionalmente na computação, enquanto que a Web Semântica é direcionada à inferência, prova e confiança. A Grade na terceira geração será construída nos termos da Grade Semântica, assim como, a Web Semântica para web, conforme mostrado na FIG. 3.1.

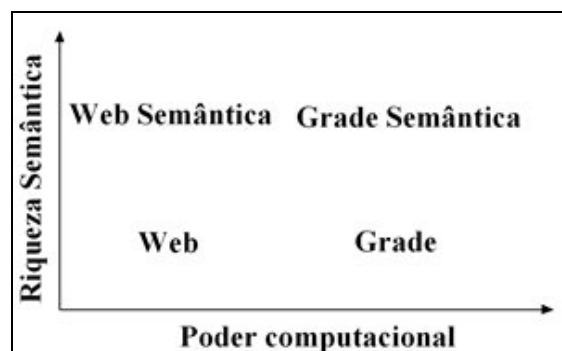


FIG. 3.1 Evolução da Grade (DE ROURE et al, 2003)

A combinação das Grades Computacionais e as tecnologias da Web Semântica já são abordadas em alguns projetos como: CombeChem (COMBECHEM), SIMDAT (ANGELE & WEITEN, 2005), Akogrimo (AKOGRIMO), ONTOGRID (ONTOGRID) e outros.

3.2 SISTEMAS P2P

P2P é uma classe de sistemas ou aplicações que se organizam automaticamente aproveitando os recursos distribuídos como: armazenamento, processamento, informação e a participação humana disponíveis na Internet (TALIA & TRIUNFIO, 2003).

A expressão ponto-a-ponto (P2P) se refere à classe de sistemas que utilizam recursos, que podem ser: poder computacional, dados (armazenamento e conteúdo), largura de banda de rede e outros, distribuídos visando realizar atividades de forma descentralizada, ou seja, sem controle centralizado ou organização hierárquica (MILOJICIC et al., 2002).

Foram desenvolvidos inicialmente para o compartilhamento de dados não-estruturados (NAPSTER), a exemplo de arquivos texto, música e vídeo, e hoje em dia, estes sistemas procuram oferecer suporte a dados estruturados e semi-estruturados, sendo chamados de *Peer Data Management System (PDMS)*.

Os PDMSs tratam de dados semanticamente mais ricos (e.g., documentos XML, tabelas relacionais, etc.) e, em sua maioria, estão focados na gerência de esquemas e na otimização e processamento de consultas. Necessitam de funções similares às dos sistemas de banco de dados distribuídos, sendo que poucos trabalhos tratam de replicação, transações e controle de acesso nesses sistemas. (VALDURIEZ & PACITTI, 2004)

Algumas características relacionadas às tecnologias P2P são descritas a seguir (MILOJICIC et al., 2002):

- *descentralização* - não utilizam um servidor central, evitando com isto, gargalo para o desempenho computacional e atualização da informação. Além disso, a execução e o custo de hardware são diminuídos;
- *escalabilidade* - está relacionada ao crescimento do volume de dados, número de conexões, capacidade da rede, número de aplicações e capacidade de processamento. Um dos benefícios da descentralização é melhorar a escalabilidade, que é limitada por

fatores como a quantidade de operações centralizadas (e.g. sincronização e coordenação dos pontos) necessárias para o funcionamento de determinados sistemas;

- *anonimato* - uma idéia do sistema P2P é permitir que as pessoas utilizem o sistema sem a preocupação de responsabilidades legais (jurídicas) e sem censura das informações digitais;
- *auto-organização* - é definida como um processo de organização de um sistema que aumenta espontaneamente, isto é, sem que este aumento seja controlado pelo ambiente ou outros sistemas externos;
- *conectividade ad-hoc* - os pontos podem se juntar ou deixar a rede de maneira *ad hoc*;
- *desempenho* - os sistemas P2P têm como objetivo melhorar o desempenho, agregando capacidade de armazenamento distribuída (e.g., Napster, Gnutella) e poder computacional (e.g., SETI@home) das máquinas dispostas na rede;
- *Tolerância a falhas* - uma das principais idéias dos sistemas P2P é evitar um ponto central de falhas. Embora alguns sistemas P2P (puro P2P) já envolvam esta idéia, os mesmo possuem falhas associadas às desconexões, inalcançabilidade, falhas dos nós, etc; e
- *Segurança* - os sistemas P2P herdaram a maioria das características de segurança dos sistemas distribuídos tradicionais, como; confiança entre os pontos e os objetos compartilhados, criptografia, assinaturas, controle de acesso e outras, sendo que novos requisitos aparecem nos sistemas P2P tais como: chaves múltiplas criptografadas, gerência dos direitos autorais, reputação e responsabilidade e *firewalls*.

Dependendo da arquitetura das redes P2P, as características acima podem ter algumas dificuldades para serem alcançadas. Estas arquiteturas independem da arquitetura física da rede e são implementadas criando-se uma rede *overlay* para os aplicativos utilizarem, estando divididas em (ANDROUTSELLIS-THEOTOKIS & SPINELLIS, 2004):

- *Pura* - todos os pontos da rede realizam as mesmas tarefas, atuando com servidores ou clientes e não existindo um ponto central para coordenação de suas atividades (FIG. 3.2);

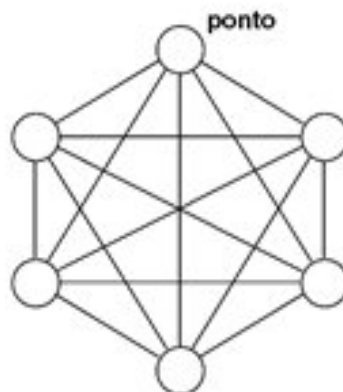


FIG. 3.2 Arquitetura Pura

- *Parcialmente centralizada* - é semelhante à topologia pura, sendo que alguns pontos assumem outros papéis importantes (ex. centralizadores de índices) e são chamados de super-pontos. As funções determinadas para os super-pontos variam entre os diferentes sistemas. Os super-pontos não constituem um único ponto de falha para a rede ponto-a-ponto em virtude de serem designados como super-ponto de maneira dinâmica e substituídos automaticamente (FIG. 3.3); e

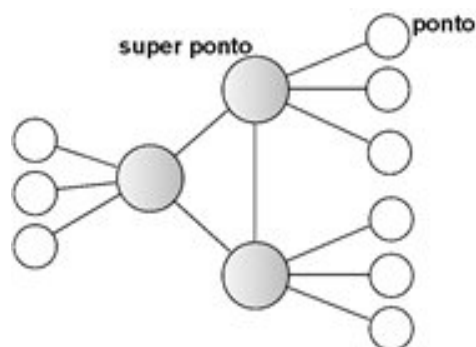


FIG. 3.3 Arquitetura Parcialmente Centralizada

- *Híbrida* - existe um servidor central que facilita a interação entre os pontos mantendo, por exemplo, diretórios de metadados e a descrição dos objetos compartilhados pelos pontos. Embora a troca de informações e a interação sejam realizadas entre dois pontos, o ponto central facilita esta interação, realizando buscas e identificando os pontos que possuem a informação solicitada por outro ponto (FIG. 3.4).

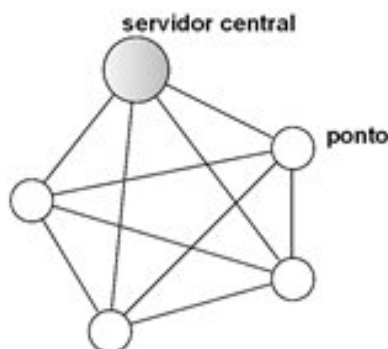


FIG. 3.4 Arquitetura Híbrida

Apesar de se apresentar como solução para muitos problemas, o alto grau de distribuição dos sistemas P2P é também a causa de alguns problemas, tais como: falta de um único esquema coerente para a organização das fontes de informação; atraso na formulação de consulta e na descoberta dos recursos; respostas duplicadas para uma única consulta e; respostas para uma única consulta geralmente requerem a integração das informações residentes em diferentes pontos independentes.

A comunidade de pesquisa tem se voltado recentemente para o uso de tecnologias da Web Semântica (ontologias, metadados e outras) nos sistemas P2P com o objetivo de reduzir estes problemas.

Alguns sistemas P2P, como os PDMSs discutidos e comparados em (FERNANDES, 2007) e (PIRES, 2007) utilizam palavras chaves, vocabulários controlados e metadados (ARENAS et al., 2003), (VALDURIEZ & PACITTI, 2004), (NG et al, 2003), (BELLAHSÈNE et al, 2004), (HALEVY et al, 2003), (OOI et al, 2003) e (BRITO & MOURA, 2005), visando adicionar relacionamentos semânticos entre as diversas fontes de dados dos pontos, trazendo assim, benefícios para o processamento das consultas.

A utilização de ontologias é reconhecida como uma abordagem efetiva para resolver problemas de heterogeneidade a nível semântico. Neste contexto, surgem os chamados sistemas de gerenciamento de dados P2P baseados em ontologias (OPDMS), os quais usam o mapeamento de ontologia para o processamento de consultas (XIAO & CRUZ, 2006.).

Nas seções seguintes (3.2.1 a 3.2.5) serão apresentados alguns trabalhos que se baseiam em descrições semânticas para a manipulação dos dados em ambientes P2P.

3.2.1 BIBSTER (BROEKSTRA et al, 2004)

O Bibster tem como objetivo compartilhar informações bibliográficas. É utilizado por pesquisadores da área de Ciência da Computação em virtude de armazenarem metadados bibliográficos no formato *BibTex*. Muitos pesquisadores possuem vários *Megabytes* de informações bibliográficas em vários arquivos *BibTex*, mas não possuem mecanismos que visam facilitar a busca, pesquisa e compartilhamento destas informações.

Os dados bibliográficos são estruturados e classificados usando duas diferentes ontologias: *Semantic Web Research Community Ontology* (SURE et al., 2005), como ontologia de aplicação e *ACM Topic Hierarchy* como ontologia de domínio (ATH). Estas ontologias são utilizadas para descrever propriedades das entradas bibliográficas no Bibster, sendo que este sistema possui uma forte dependência da ontologia de aplicação e a ontologia de domínio pode ser facilmente substituída para adaptação a diferentes domínios. Estas ontologias são utilizadas também para ajudar os usuários a formularem suas perguntas, para melhorar o roteamento na rede P2P e para processar as respostas detectando ambigüidades.

O Bibster utiliza a *API BibToOnto* para transformar, automaticamente, a representação dos metadados bibliográficos dos arquivos *BibTex* em uma representação baseada em ontologias (RDF) e também para a classificação das entradas bibliográficas de acordo como a ontologia *ACM topic hierarchy*.

A FIG. 3.5 apresenta uma visão geral da arquitetura do Bibster, onde cada ponto participante da rede possui os mesmos componentes, descritos a seguir:

- *Communication Adapter* - o adaptador de comunicação é responsável pela comunicação da rede entre os pontos. Atua como uma camada de transporte para outras partes do sistema, isto é, envio e reenvio de consultas. A plataforma *JXTA* é responsável por esta funcionalidade;
- *Knowledge Sources* - as fontes do conhecimento são fontes de metadados bibliográficos, i.e., arquivos *BibTex* armazenados localmente no sistema de arquivos;
- *Knowledge Source Integrator* - o integrador da fonte de conhecimento é responsável pela extração e integração de fontes internas e externas de conhecimento para o *Local Node Repository*;
- *Local Node Repository* - é baseado no *RDF-S Repository Sesame* e as linguagens *SeRQL*, *RQL* ou *RDQL* são utilizadas para realizar consultas semânticas. O repositório

possui as seguintes funcionalidades: armazenar e prover visões do conhecimento disponível; dar suporte a operações como formulação de consultas e processamento sobre o conhecimento disponível; especificação de uma interface do ponto para a rede; e fornecimento de uma base para a classificação e a seleção do ponto;

- *Informer* - tem com função publicar o conhecimento disponível no ponto para rede e descobrir pontos com informações relevantes para as respostas das consultas do usuário. Isto é realizado emitindo anúncios sobre a perícia do ponto. No sistema Bibster, as descrições do ponto ficam contidas num conjunto de tópicos. Os pontos armazenam estes anúncios e podem utilizá-los para criação de ligações semânticas com outros pontos. Estas ligações dão forma a uma topologia semântica, que é a base para o roteamento inteligente da consulta;
- *Query Replier* - coordena a distribuição das consultas. Recebe consultas da interface do usuário e as distribui de acordo com o conteúdo da pergunta. Quando o ponto recebe uma pergunta de outro ponto, tenta responder ou reenviar para outro ponto. A decisão de/para quais pontos uma consulta deva ser enviada é feita com base no conhecimento sobre a perícia de outros pontos;
- *User Interface* - A interface do usuário permite a importação, criação e edição de metadados bibliográficos, bem como formular consultas de maneira intuitiva. O usuário pode também formular consultas semânticas avançadas com base na ontologia *SWRC* e na hierarquia do tópico *ACM*. Além disso, a consulta poderá também ser direcionada ao próprio ponto, a um determinado ponto ou a um grupo de pontos. Os resultados poderão ser visualizados e integrados no repositório local, ou exportados em BibTeX ou HTML.

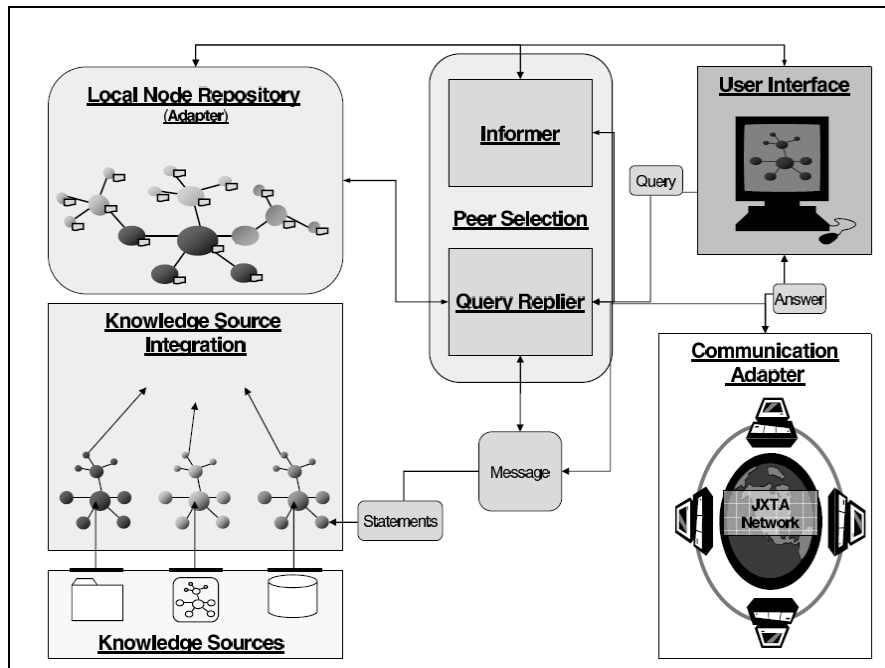


FIG. 3.5 Arquitetura do Bibster (BROEKSTRA et al, 2004)

No Bibster somente uma ontologia é utilizada para troca de informações, obrigando os pontos a organizarem suas informações segundo esta ontologia. Dessa forma, apesar de funcionarem de forma descentralizada e auto-organizável, a atualização de ontologia, em um dado momento, depende de um ponto central, sendo assim a descentralização e auto-organização são comprometidas.

3.2.2 OYSTER (PALMA et al, 2005)

Oyster é uma aplicação P2P que explora técnicas da Web Semântica a fim de fornecer uma solução para troca e reuso de ontologias. Tem como objetivo auxiliar pesquisadores na gerência, busca e compartilhamento de metadados de ontologias em um ambiente P2P.

O sistema usa um padrão de metadados que descreve as propriedades da ontologia, chamada de *Ontology Metadata Vocabulary - OMV* (OMV), onde informações são instanciadas e, também uma hierarquia de tópicos denominada *DMOZ* (DMOZ), que descreve categorias específicas de assuntos, visando definir o domínio da ontologia.

O uso dessas ontologias visa auxiliar as principais funções do sistema, descritas a seguir:

- *Criação e importação de metadados* - permite que usuários criem metadados sobre ontologias manualmente, podendo também importar arquivos de ontologia a fim de extrair automaticamente os metadados das ontologias e preencher outros dados manualmente, caso não estejam disponíveis na ontologia. A entrada dos metadados da ontologia são alinhadas e formalmente representadas de acordo com a *OMV* e hierarquia de tópicos *DMOZ*;
- *Formulação de consultas* - usuários podem procurar ontologias no repositório por meio de buscas de palavras chaves simples, ou buscas mais avançadas, envolvendo características semânticas. As consultas são formuladas em termos das duas ontologias citadas anteriormente. Isto significa que as consultas podem incluir campos como nome, acrônimo, linguagem da ontologia, etc., ou podem consultar por termos da categoria da ontologia;
- *Roteamento de consultas* - usuários podem consultar um único ponto específico (por exemplo, seu próprio ponto ou outro conhecido por conter grande quantidade de informação), um conjunto específico de pontos (por exemplo, todos os membros de uma organização específica) ou a rede inteira (por exemplo, quando o usuário não tem nenhuma idéia onde procurar). Neste último caso, as consultas são roteadas automaticamente através da rede, fazendo uso da especialidade atribuída aos pontos, cujos tópicos são definidos na *DMOZ*. Uma função de similaridade é utilizada para determinar se o conteúdo semântico da consulta e a especialidade do ponto são compatíveis;
- *Processamento de resultados* - os resultados de uma consulta são apresentados em uma lista. Estes resultados podem ser muito grandes e/ou conterem muitas duplicatas devido à natureza distribuída e ao tamanho potencialmente grande de uma rede P2P. Tais duplicatas podem não ser cópias exatas por causa da natureza semi-estruturada dos metadados. Sendo assim, as ontologias são usadas novamente para medir a similaridade semântica entre as diferentes respostas, removendo as duplicatas evidentes. Por fim, os usuários podem salvar os resultados de uma consulta em seu repositório local para uso futuro.

O Oyster é semelhante ao Bibster, onde somente uma ontologia é utilizada para troca de informações, obrigando os pontos a organizarem suas informações segundo esta ontologia. Dessa forma, apesar de funcionarem de forma descentralizada e auto-organizável, a

atualização de ontologia depende de um ponto central, sendo assim a descentralização e a auto-organização, também, são comprometidas.

3.2.3 ROSA - P2P (BRITO & MOURA, 2005)

O sistema ROSA-P2P tem como objetivo viabilizar a integração de objetos de aprendizagem - Learning Objects (LOs) em um ambiente distribuído, onde usuários podem submeter consultas e receber os respectivos resultados, resultantes da integração dos dados residentes nos diversos pontos ROSA - P2P.

Sua arquitetura é baseada em super-pontos, e inclui estratégias específicas para: conexão/desconexão de pontos à rede P2P, tais como o agrupamento de super-ponto por assunto; definição e eleição de super-pontos; balanceamento e redistribuição de pontos no sistema; e alguns aspectos de tolerância à falhas.

O ROSA-P2P possui um sistema de integração de dados P2P que utiliza estruturas semânticas, denominadas de vocabulários controlados, e possui mecanismos para o correto reenvio, reescrita e execução de consultas, assim como para integração dos respectivos resultados. Sendo assim, cada consulta é reenviada aos pontos relevantes, os quais as reescrevem com base na semântica do seu domínio e as executam com base na álgebra ROSA. Os resultados parciais, oriundos de cada ponto ROSA - P2P, são enviados ao ponto solicitador, que os integra e os retorna ao usuário.

A arquitetura interna do ROSA-P2P possui três módulos, conforme, apresentados na FIG. 3.6, a saber (BRITO, 2005):

- *Módulo de interoperabilidade*: formado pelo componente interoperador (P2P), possui as características e funcionalidades necessárias para a formação e manutenção de uma rede P2P, tais como estabelecimento de conexões, atualizações de índices de roteamento, eleição de super-pontos e balanceamento da rede;
- *Módulo de processamento de consulta*: formado pelo componente interface do usuário, é responsável por oferecer um ambiente mais amigável de comunicação entre os usuários e o componente de processamento de consultas, que adota uma estratégia dividida em 2 fases

- *Módulo de gerenciamento de dados*: é formado por dois componentes: *vocabulários controlados*, que são ferramentas que facilitam e possibilitam a interpretação e recuperação das informações, ao mesmo tempo em que viabilizam o intercâmbio entre os sistemas, permitindo pesquisas mais apuradas e restritas às informações realmente relevantes; e *cache/integrador de dados*, que armazena temporariamente os resultados parciais das consultas. Uma vez obtido os resultados dos pontos e/ou super-pontos para os quais a consulta foi reenviada, é realizada sua integração e o resultado final da consulta é exibido ao usuário.

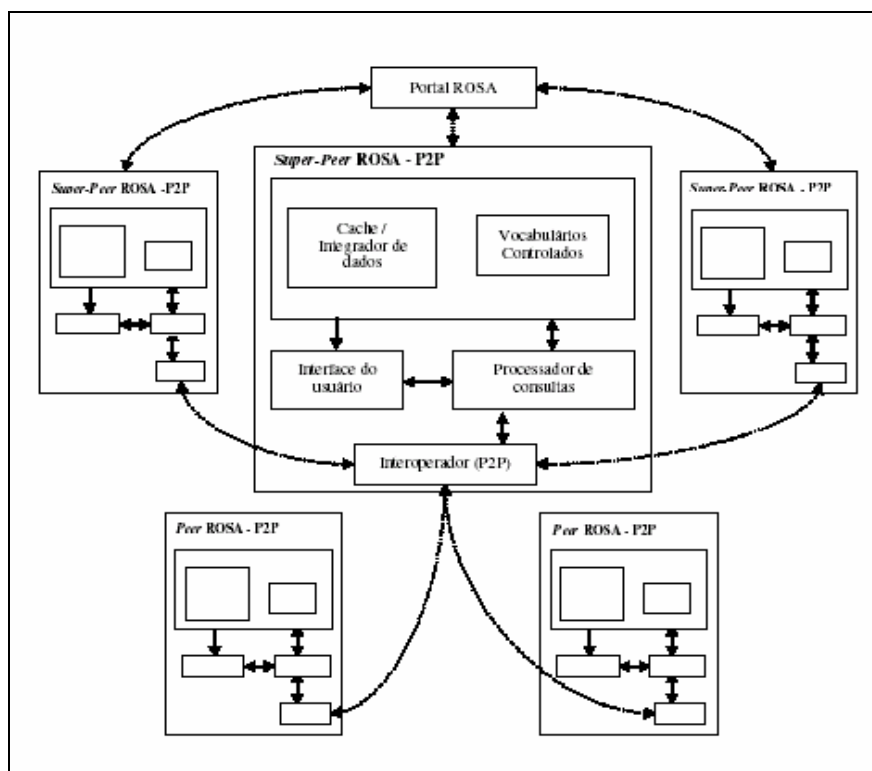


FIG. 3.6 Arquitetura do ROSA-P2P (BRITO, 2005)

O ROSA-P2P utiliza vocabulários controlados (global, local e palavras-chaves) facilitando a reescrita de consultas. Estes vocabulários, apesar de persistidos nos pontos, sua obtenção, depende de uma ponto central (Portal Rosa) que não considera a idéia de gerar e/ou atualizar de forma automática estes vocabulários com base nas reescritas de consultas realizadas pelos diversos pontos.

3.2.4 XPEER (BELLAHSÈNE et al., 2004)

O XPeer é um PDMS que permite consultas em grande escala, abstraindo da distribuição, localização e heterogeneidade das fontes de dados e proporcionando tempos de resposta mais aceitáveis.

A arquitetura do sistema é baseada no conceito de super-pontos, organizados em redes lógicas baseadas na semântica dos dados, os quais direcionam o processamento de consultas e troca de dados entre os pontos.

As fontes de dados dos pontos são organizadas seguindo três regras, a saber:

- **Grupo:** os pontos são organizados em grupos, onde um determinado ponto possui um esquema do grupo que realiza a mediação entre as fontes de dados dos pontos participantes. Cada fonte de dados dos pontos é definida como uma visão sobre o esquema do grupo usando regras de mapeamento *local-as-view* (LAV). O grupo possui mapeamentos diretos entre as fontes de dados dos pontos;
- **Domínio:** um domínio reúne grupos que compartilham a mesma categoria de informação. Fornece um único serviço de diretórios sobre os grupos, que contém um conjunto de palavras chaves fornecidas pelos grupos no momento que publicam seu esquema no domínio;
- **Global:** é o ponto de entrada no sistema onde se pode localizar o domínio correspondente ao ponto de interesse. O ponto global registra um conjunto de palavras chaves que identificam as categorias das informações que são abrangidas pelos domínios.

A estrutura formada por estes tipos de pontos reduz a quantidade de mapeamentos entre as fontes de dados dos pontos, aumentando, dessa forma, a escalabilidade da arquitetura. Diante disso, os pontos de domínio e o ponto global são integrados no mais alto nível de abstração, fazendo uso de palavras chaves. Os pontos de grupo e as fontes de dados dos pontos são integrados em um nível mais baixo, usando padrões de mapeamentos de dados. Sendo assim, as fontes de dados dos pontos podem sair e entrar na rede P2P facilmente, afetando somente o ponto de grupo, tornando a arquitetura mais flexível.

No caso de ter um ponto global, um único nível de pontos de domínio e um único domínio para os pontos de grupo (FIG. 3.7) o sistema torna-se propenso a falhas. Para resolver este problema o XPeer foi estendido para o iXPeer.

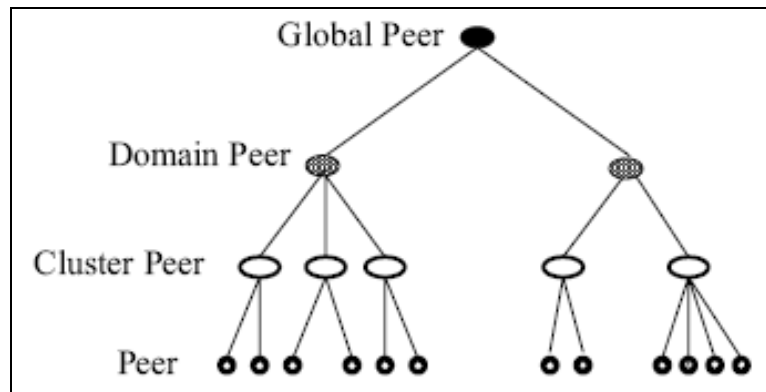


FIG. 3.7 Organização da fontes de dados do XPeer (BELLAHSÈNE et al, 2004)

A extensão da arquitetura do XPeer para iXPeer (BELLAHSÈNE et al, 2006) foi implementada utilizando as características principais do AutoMed P2P (BOYD et al, 2004), dispoñdo de uma hierarquia mais flexível, com um número maior de níveis de domínio e habilidade de reunir múltiplas hierarquias. Na FIG. 3.8 é apresentada a nova implementação, sendo que não existe distinção entre o ponto global e de domínio.

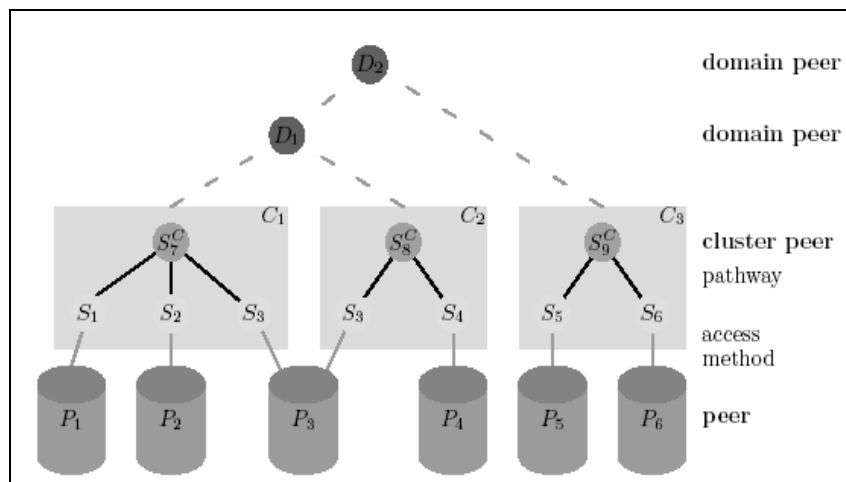


FIG. 3.8 Extensão do XPeer (BELLAHSÈNE et al, 2006)

O XPeer estrutura os pontos em grupos, domínio e global, com a finalidade de reduzir os mapeamentos entre os pontos, formando uma hierarquia, o que aumenta a escalabilidade. Apesar de não ser uma arquitetura descentralizada, possui mecanismos de tolerância a falhas que caso um ponto esteja fora da rede, o sistema se auto-organiza.

3.2.5 SPEED (DAMIRES, 2007) E (PIRES, 2007)

O Sistema P2P de gerenciamento de dados baseado em semântica (*Semantic PEER-to-peer Data Management System – SPEED*) é um projeto de pesquisa que tem como objetivo desenvolver soluções para resolver problemas de conectividade, mapeamento entre esquemas, processamento de consultas e qualidade de serviços no que tange ao gerenciamento de dados em uma rede P2P. Utiliza abordagens baseadas em semântica como principal artefato para auxiliar na resolução dos problemas encontrados no gerenciamento de dados P2P.

A arquitetura do SPEED, apresentada na FIG. 3.9, é composta por três tipos de pontos: semânticos, integração e dados. Utiliza duas topologias de redes distintas: a rede DHT que organiza os pontos semânticos e a rede super-pontos que gerencia os pontos de integração e de dados.

Os pontos semânticos servem de ponto de entrada nas comunidades semânticas, realizam o balanceamento dos pontos e resolvem onde um ponto que deseja se conectar deve ser conectado. Atuam como servidores de ontologias específicas de um determinado domínio. Estas ontologias são utilizadas para enriquecer semanticamente os esquemas exportados dos pontos e distribuí-los, de modo eficiente dentro das comunidades semânticas e dos grupos (*clusters*).

Os pontos de integração são os pontos de maior poder computacional dentro de um determinado grupo semântico, em virtude destes realizarem todo o controle e processamento das consultas aos pontos a ele conectados. Eles se comunicam com outros pontos de integração e realizam a comunicação dos pontos de dados, a ele conectados, com o restante da rede de forma eficiente.

Os pontos de dados são fontes que podem ser compartilhadas com os demais pontos através de mapeamentos semânticos, podendo também servir de local de execução das consultas.

Os grupos semânticos associam logicamente um ponto de integração com um determinado conjunto de pontos de dados que possuem relações semânticas comuns. O ponto de integração de dentro do grupo tem como funções realizar a indexação de metadados, o processamento de consultas e a integração dos dados.

As comunidades semânticas associam logicamente grupos semânticos que possuem interesses semânticos comuns. O domínio semântico basicamente define a organização dos

pontos, auxiliando no processo de entrada dos pontos na rede, troca de informações, balanceamento de carga dos grupos semânticos e tolerância a falhas.

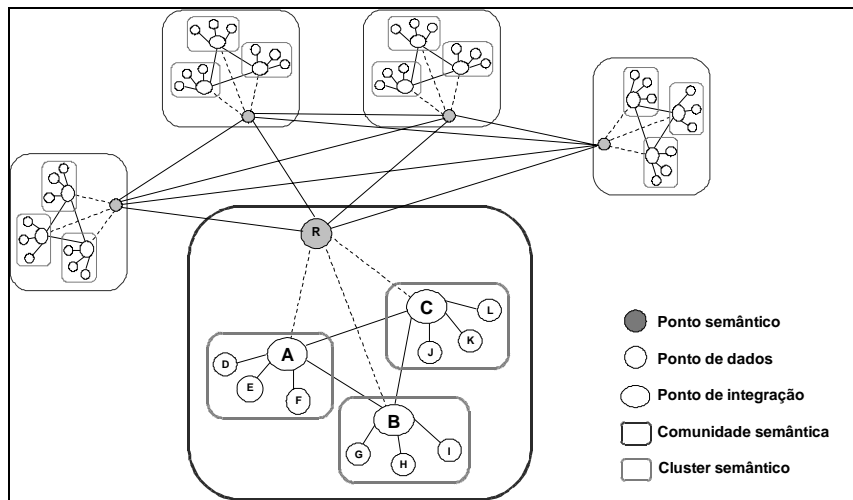


FIG. 3.9 Arquitetura geral do sistema SPEED (PIRES, 2007)

O SPEED possui uma ontologia padrão de um domínio pré-definida no ponto semântico, com foi fito anteriormente, uma meta-ontologia no ponto de integração chamada de Ontologia de Referência do *Cluster (ORC)*, onde são definidos os mapeamentos semânticos entre os pontos de dados e de integração e uma meta-ontologia contextual, onde são instanciados os dados contextuais. Pode-se observar que esta ferramenta não leva em conta os mapeamentos da *ORC* para auxiliar a geração e/ou o enriquecimento da ontologia de domínio.

3.3 GRADE X P2P

Esta seção tem como principal finalidade analisar e comparar as características principais de sistemas P2P e Grades, baseando-se em (TALIA & TRIUNFO, 2003).

- **Segurança**

Grade - Implementada em comunidades fechadas. Por natureza não permitem o anonimato dos usuários e nem dos recursos. Existem pesquisas visando explorar modelos de segurança dos sistemas P2P.

Sistemas P2P - Os sistemas são implementados em comunidades abertas, nas quais os usuários podem compartilhar qualquer informação. Por este motivo os mecanismos de segurança na maioria dos sistemas P2P não realizam a autenticação e validação do conteúdo, sendo assim não há restrições quanto ao anonimato e censura.

- **Conectividade**

Grade - Em geral usam máquinas de alto desempenho que são conectadas que forma estática em redes de alto desempenho e com alto nível de disponibilidade. O número de nós participantes geralmente são menores devido ser limitado por rigorosos mecanismos de acesso, sendo que poderiam se beneficiar da flexibilidade dos modelos de conectividade usados nos sistemas P2P.

Sistemas P2P - São compostos geralmente por máquinas de trabalho que se conectam de modo dinâmico na rede, permanecendo disponível por tempo limitado, com segurança reduzida.

- **Acesso aos serviços**

Grade - A idéia primária de acesso aos recursos remotamente na construção das Grades ainda permanece, hoje em dia. Algumas ferramentas proporcionam serviços com segurança, como computação em lote ou executando aplicações interativas em máquinas remotas. Possuem mecanismos para o compartilhamento, armazenamento e movimentação de dados entre os nós.

Sistemas P2P - Não suportam mecanismos que alocam explicitamente e remotamente o local de processamento ou armazenamento. Possuem mecanismos para o compartilhamento e troca de dados entre os nós.

- **Descoberta de recursos e gerenciamento**

Grade - é baseada principalmente em modelos centralizados ou hierárquicos. As informações sobre os recursos disponíveis em determinado nó podem ser consultadas diretamente nele ou em outro nó onde são publicados informações de um conjunto de nós. Não trabalham em ambientes muito dinâmicos e altamente distribuídos, onde as informações necessárias dos servidores não são previamente conhecidas. Em parte, inclui questões de gerência e para descoberta dos nós que estão disponíveis na Grade, devido não possuir mecanismos. Futuramente as Grades deverão ser implementadas em modelos de descoberta de recursos de forma descentralizada, semelhante aos modelos P2P, apoiando a idéia de Grades em comunidades abertas.

Sistemas P2P - questões relacionadas a protocolos de gerência dos nós são elementos importantes em sistemas P2P. Os nós periodicamente notificam sua presença na rede e descobrindo seus vizinhos ao mesmo tempo.

- **Tolerância à falhas**

Grade - Sabe-se que muitas aplicações executam tarefas em diversos nós da rede, sendo assim, necessitam de um alto nível de tolerância à falhas. Possui mecanismos de *checkpointing* e de inicialização, sendo que segurança e tolerância a falhas não são muito explorados em seus modelos e ferramentas.

Sistemas P2P - Evitam serviços centralizados que podem representar pontos críticos de falhas. São muito explorados em seus modelos e ferramentas.

Ambas as abordagens poderiam se aplicar ao problema de integração de informações em uma grande organização, mas P2P parece atender ao problema de modo mais fácil, sem a sofisticação que caracteriza os ambientes de Grade, e ainda não há necessidade de compartilhar recursos para o processamento, que é o foco da abordagem de Grade.

Além disso, segundo (TALIA & TRIUNFIO, 2003) os modelos P2P podem ser usados nas próprias Grades, assegurando a escalabilidade e a implementando Grades não hierárquicas e descentralizadas.

Dos sistemas P2P podemos destacar o seguinte:

- não necessitam computadores de alto desempenho;
- permitem flexibilidade no armazenamento e gerenciamento de informações, facilitando o seu reuso e compartilhamento entre os diversos pontos;
- permitem maior flexibilidade no processo de mapeamentos entre esquemas, com o uso de metadados e ontologias, melhorando e facilitando significativamente: o entendimento, a resolução dos possíveis conflitos e a integração dos dados residentes nos diversos pontos;
- a conectividade dos pontos, apesar de instáveis, comparado com as Grades, não limita a abordagem proposta; e
- não necessita de altos níveis de tolerância a falhas.

3.4 CONSIDERAÇÕES FINAIS

A idéia principal deste capítulo foi apresentar as principais abordagens sobre os sistemas de integração de dados, com enfoque para as estruturas P2P e Grades computacionais.

Foi possível identificar que as abordagens de integração de dados P2P, não têm como principal função gerar uma visão global de forma automática fazendo uso dos mapeamentos estabelecidos entre os pontos ou sítios. O Bibster e Oyster utilizam ontologias para padronizadas para organização dos dados e o processamento das consultas. O XPeer não utiliza ontologias. Os esquemas dos pontos são exportados e são realizados mapeamentos para um esquema global, com o intuito de melhorar as consultas. O ROSA-P2P faz uso de vocabulários controlados pré-definidos e entregues aos pontos no momento entram na rede ou quando os seus dados são atualizados e o SPEED faz uso de informações de contexto, domínio e dos mapeamentos entre os pontos, mas não são utilizados para geração de uma visão global.

Após análises dos principais aspectos relacionados a integração de dados em ambientes autônomos e disponibilização de uma ferramenta para desenvolvimento, o COPPEER (MIRANDA, 2006), que será apresentada no capítulo 5, optou-se pelos sistemas P2P para o desenvolvimento do própotipo.

O próximo capítulo visa explorar a arquitetura da abordagem Ontologias Emergentes, apresentando as funcionalidades dos módulos e bases de dados necessários para gerar uma visão global, bem como definições (básicas e de constantes) e heurísticas aplicadas aos mapeamentos realizados entre os diversos pontos de uma rede.

4 ONTOLOGIAS EMERGENTES

Este capítulo apresenta a proposta de Ontologias Emergentes para a geração de uma ontologia. Esta proposta é uma nova abordagem para a integração de ontologias, que busca agilizar a geração de ontologias corporativas. De fato, abordagens propostas até o momento para integração (FARQUHAR et al., 1996) e construção de ontologias (NOY, et al., 2001), (USCHOLD, et al., 1995), (GUARINO, 1997), (FERNANDEZ, et al., 1997), (GRUBER, 1995) e (GRÜNINGER & FOX, 1995) não atendem de maneira efetiva este objetivo, já que foca em soluções voltadas para integração de um par de ontologias ou em soluções voltadas para metodologias de construção de uma Ontologia, a partir de várias. Essas abordagens não consideram a existência de várias ontologias.

A abordagem de Ontologias Emergentes parte dos mapeamentos entre ontologias disponíveis em um ambiente distribuído, como ocorre, por exemplo, em uma grande corporação, ou até mesmo na Web, visando a geração de uma visão global, a qual é chamada, no contexto desta dissertação, de Ontologia Emergente (OE). Em um ambiente é comum que hajam iniciativas de construção de ontologias setoriais, que atendam a contextos menores, como um departamento, projeto ou sistema. Além disso, para um determinado domínio, pode haver iniciativas de mapeamento entre tais ontologias, interligando diferentes "pontos" do ambiente distribuído. Isso se dá na medida em que um ponto deseja trocar informações sobre um dado domínio, com outros pontos, sendo que um mapeamento entre as respectivas ontologias é realizado, e uma representação deste mapeamento é gerada. Os diversos pontos da organização acumulam estes mapeamentos ponto-a-ponto ao longo do tempo. A abordagem de Ontologias Emergentes pressupõe que no momento em que a quantidade de mapeamentos atinge um determinado número, é possível então, a partir de uma análise sobre os conceitos das diversas ontologias e seus mapeamentos, gerar uma ontologia emergente, ou visão global da corporação. Assim, tem-se um ambiente onde as ontologias estão distribuídas pelos diversos "sítios" ou "pontos" e os mapeamentos entre estas ontologias interligam estes pontos, caracterizando-o como um ambiente P2P.

As arquiteturas distribuídas P2P são caracterizadas pelo compartilhamento de recursos computacionais e serviços através da comunicação direta e descentralizada entre os sistemas

(pontos) envolvidos (BRITO, 2005). Os recursos e serviços oferecidos por estas arquiteturas incluem mecanismos que facilitam a troca de informações. Algumas abordagens deste tipo de arquitetura (BROEKSTRA et al, 2004) e (PALMA et al, 2005) utilizam ontologias para facilitar esta troca de informações. Entretanto, embora estas abordagens ofereçam um modo descentralizado de representação da informação, costumam considerar ontologias similares, ou até mesmo idênticas, residentes em cada ponto. Além disso, não possuem como objetivo a geração de uma ontologia representativa de todos os pontos.

Como o foco deste trabalho é a geração de uma visão global (Ontologia Emergente) baseada na troca de informações entre pontos de uma organização, como ocorre nos sistemas P2P, propõe-se então a extensão de uma arquitetura P2P para atingir este objetivo.

Para o entendimento do sistema proposto, este capítulo foi organizado em três seções. A seção 4.1 apresenta a arquitetura proposta, detalhando o seu funcionamento. Logo após, a seção 4.2 apresenta as definições básicas e as heurísticas usadas para a seleção dos conceitos e propriedades que passam a compor a ontologia emergente, e que serviram de base para o desenvolvimento do protótipo implementado. Por fim, a seção 4.3 faz algumas considerações finais sobre o capítulo.

4.1 ARQUITETURA

Esta seção visa apresentar e especificar de modo abrangente os componentes da arquitetura proposta. Como foi dito anteriormente, a arquitetura proposta estende uma arquitetura típica P2P visando proporcionar uma visão global da organização a partir da troca de informações entre os pontos da rede P2P. O tipo de arquitetura P2P adotado é o que provê uma estrutura pura, na qual cada ponto é equivalente em funcionalidade, e não há um controle centralizado ou organização hierárquica.

A FIG. 4.1 mostra a arquitetura da abordagem de Ontologias Emergentes. Tipicamente, um sistema P2P oferece uma infra-estrutura básica para comunicação e gerência, que é representada pelo módulo (Comunicação e Gerência P2P) que sustenta os demais módulos da arquitetura proposta. Os demais módulos foram acrescentados para alcançar os objetivos desta proposta. Estes módulos e as bases de dados associadas são descritos sucintamente em seguida e detalhados nas seções subseqüentes.

O módulo de Mapeamento P2P tem por finalidade realizar mapeamentos entre ontologias. O módulo de Gerência tem como objetivo principal verificar o momento de iniciar a geração da OE. O módulo de Geração OE verifica os mapeamentos realizados pelo módulo de Mapeamento P2P e, com o uso de heurísticas, elege os conceitos que serão representados na OE. O módulo de Navegação-Instanciação possibilita a visualização das instâncias de duas formas: local e global. O módulo de rastreamento monitora algumas atividades do sistema. Por fim, as bases de dados (BD MAP, BD ONTOLOGIAS e BD INSTÂNCIAS) são utilizadas para o armazenamento de arquivos de mapeamentos, ontologias e instâncias, respectivamente.

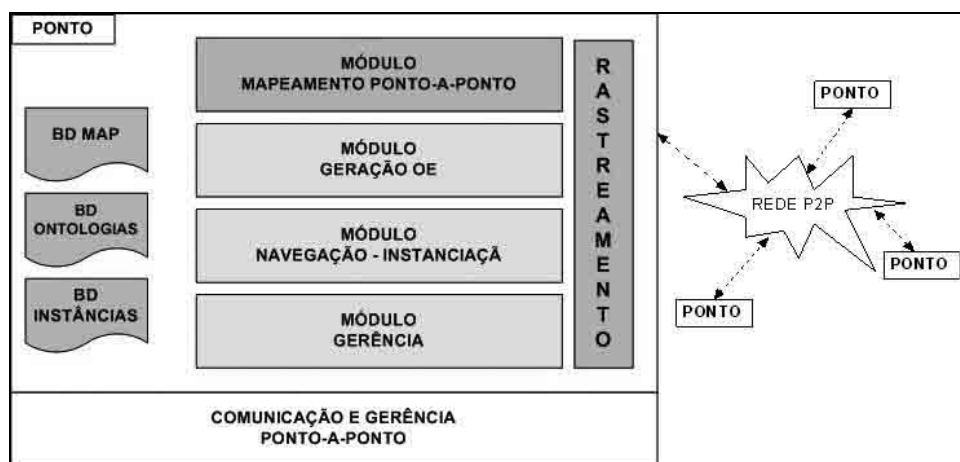


FIG. 4.1 Arquitetura Ontologias Emergentes

4.1.1 DINÂMICA DE MAPEAMENTOS E GERAÇÃO DA OE

Para o melhor entendimento da arquitetura proposta, esta seção apresenta uma dinâmica entre quatro pontos X, Y, J e Z, cada qual possuindo uma ontologia. As ontologias X, Y, J e Z pertencem, respectivamente, aos pontos X, Y, J e Z. A FIG. 4.2 mostra dez situações que acontecem entre estes quatro pontos até a geração da OE em um deles.

Na situação 1 o ponto X realiza um mapeamento com o ponto Y, gerando um arquivo de mapeamento chamado de MO_XO_Y no ponto X.

Na situação 2 ocorre a transferência do arquivo de mapeamento MO_XO_Y para o ponto Y.

Na situação 3 o ponto X realiza um mapeamento com o ponto Z, gerando um arquivo de mapeamento chamado de MO_XO_Z no ponto X.

Na situação 4 ocorre a transferência dos arquivos de mapeamentos MO_XO_Z e MO_XO_Y para o ponto Z.

Na situação 5 o ponto Z realiza um mapeamento com o ponto J, gerando um arquivo de mapeamento chamado de MO_ZO_J no ponto Z.

Na situação 6 ocorre a transferência do arquivo de mapeamento MO_ZO_J e dos arquivos de mapeamentos recebidos de outros pontos MO_XO_Z e MO_XO_Y para o ponto J.

Na situação 7 o ponto Y realiza um mapeamento com o ponto J, gerando um arquivo de mapeamento chamado de MO_YO_J no ponto Y.

Na situação 8 ocorre a transferência do arquivo de mapeamento MO_YO_J e MO_XO_Y para o ponto J.

Na situação 9, um dos pontos verifica que tem condições de gerar a OE, analisando a quantidade de mapeamentos realizados ou recebidos de outros pontos, e se estes mapeamentos envolvem todos os pontos da rede P2P. Na verdade, este monitoramento é realizado de periodicamente.

Na situação 10, uma vez que o ponto esteja apto a gerar a OE, esta poderá ser gerada e disponibilizada para os demais, proporcionando, assim, uma visão global, tendo como base os mapeamentos realizados entre os pontos.

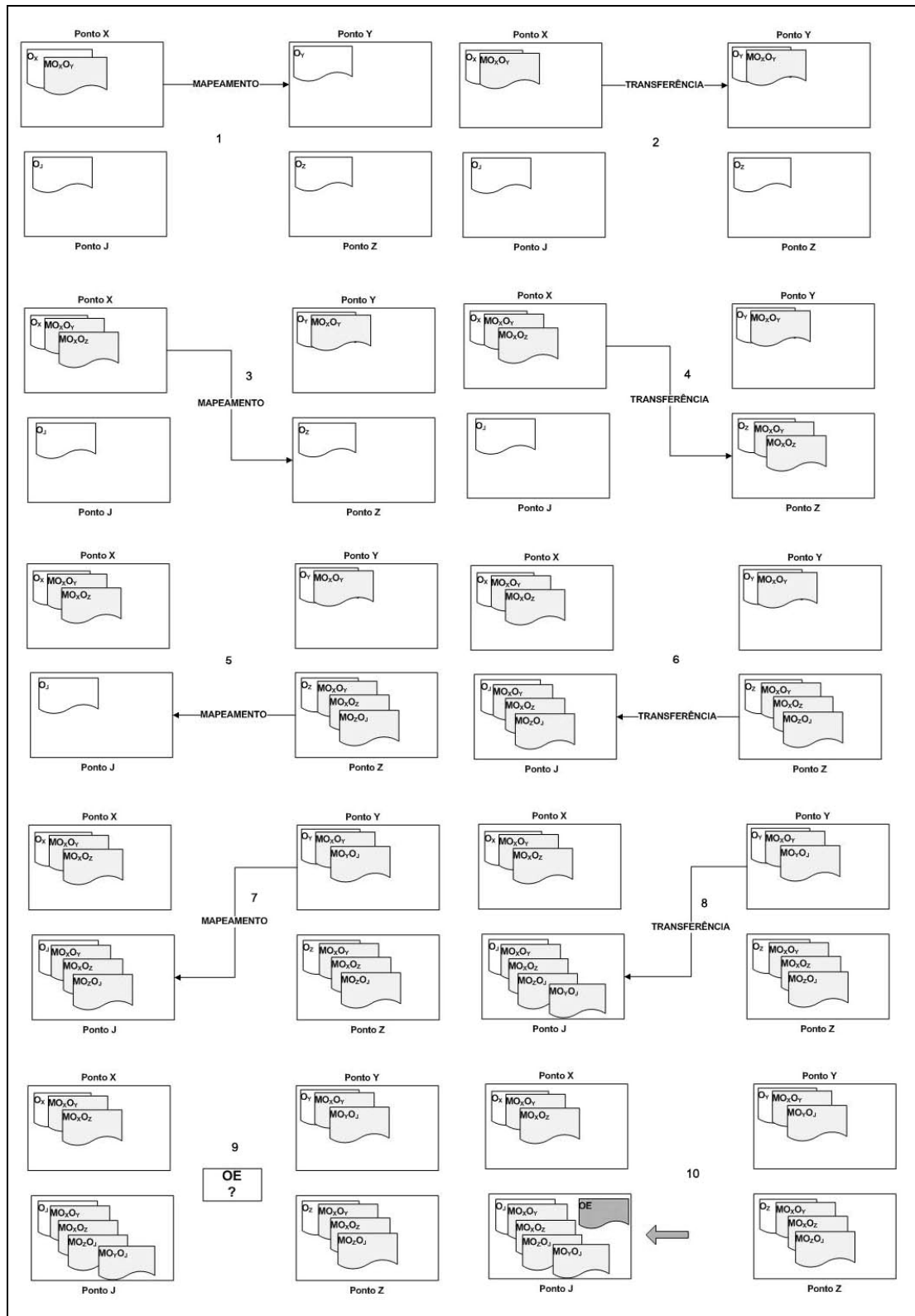


FIG. 4.2 Exemplo de dinâmica dos mapeamentos entre quatro pontos

4.1.2 MÓDULO DE MAPEAMENTO P2P

A finalidade deste módulo é estabelecer os mapeamentos entre duas ontologias de pontos distintos e armazená-los no BD MAP de ambos os pontos. Os mapeamentos poderão ser alcançados de forma manual (MAEDCHE et al., 2002) e (NUNO SILVA, 2004) ou semi-automática (FALCONER et al., 2006), normalmente utilizando ferramentas de mapeamento de ontologias. No capítulo 2 foram descritos algumas ferramentas encontradas na literatura, verificando-se que algumas abordagens de mapeamento geram uma forma de representação dos mapeamentos e mantêm, em sua forma original, as ontologias envolvidas no mapeamento. As representações destes mapeamentos servirão de base para construção da visão global da organização, isto é, da ontologia da organização.

Na FIG. 4.3 são apresentadas duas ontologias, 1 e 2, onde mapeamentos são realizados. Nela pode-se verificar, o mapeamento entre o conceito O1.DOCUMENTOS e o conceito O2.EXPEDIENTE. Estes mapeamentos são armazenados, pois serão utilizados posteriormente para a geração da OE e para realizar transformações de instâncias entre as ontologias envolvidas. Na FIG. 4.4a tem-se as instâncias das duas ontologias e na FIG. 4.4b a instância OFICIO 10/2007 do conceito O1.DOCUMENTOS transferida para o conceito O2.EXPEDIENTE. Estes arquivos de mapeamentos serão os principais artefatos a serem analisados na abordagem proposta por esta dissertação.

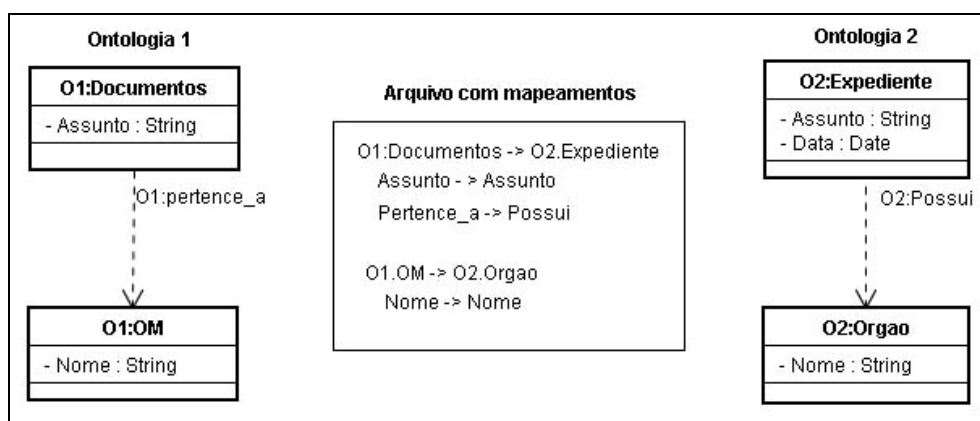


FIG. 4.3 Exemplo de mapeamentos entre as ontologias 1 e 2

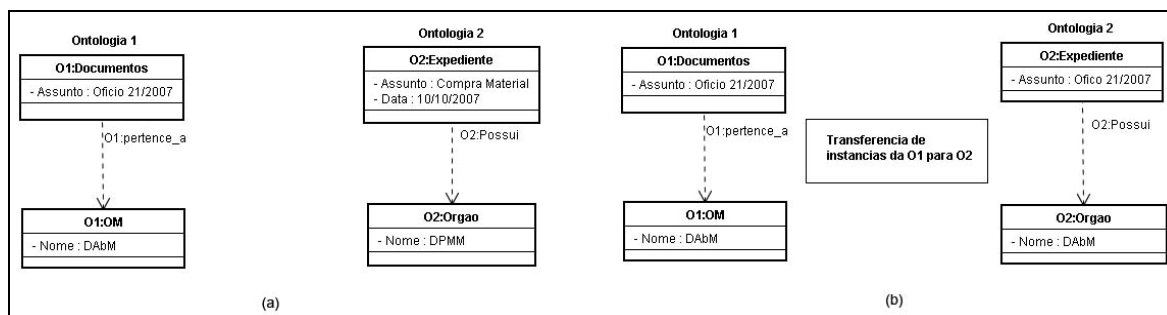


FIG. 4.4 Instâncias das ontologias 1 e 2

A base de dados (BD MAP), descrita na seção 4.1.7, armazenará os arquivos com as representações dos mapeamentos realizados entre os pontos, permitindo o acesso por outros módulos descritos ao longo deste capítulo.

4.1.3 MÓDULO DE GERÊNCIA

Este módulo monitora periodicamente (intervalo de tempo a ser definido na configuração de cada ponto) a quantidade de mapeamentos entre ontologias armazenados no BD MAP. Quando esta quantidade alcança um limite mínimo pré-estabelecido (na configuração do sistema), este módulo sinaliza ao *Módulo Geração da OE* que já existe a possibilidade de gerar a OE. Além disso, verifica e informa aos diversos pontos participantes da rede P2P, sobre a existência da OE em um determinado ponto. Outra responsabilidade deste módulo é oferecer uma interface de configuração para o usuário, de modo que possa definir, por exemplo, o intervalo de tempo de monitoramento do BD MAP, o nome do ponto, os diretórios do sistema operacional que serão utilizados, entre outros dados de configuração.

4.1.4 MÓDULO DE GERAÇÃO DA OE

Quando sinalizado pelo módulo de Gerência, informando sobre a possibilidade de gerar a OE, este módulo realiza a análise dos mapeamentos existentes no ponto e aplica as heurísticas com a finalidade de definir quais os conceitos e propriedades (atributos e relações) que poderão compor a OE.

Os mapeamentos realizados entre duas ontologias de pontos distintos são armazenados no BD MAP de cada ponto, como visto anteriormente. Estas representações de mapeamentos entre conceitos serão analisadas em nossa abordagem. Uma heurística para eleição dos conceitos que devem compor a OE será aplicada, considerando os conceitos mais mapeados, direta ou indiretamente. Um detalhamento desta heurística é apresentado na seção 4.2.3. Por exemplo, se os diversos pontos mapeiam para um determinado conceito "c", ele provavelmente será eleito e comporá a Ontologia da organização.

Uma vez definidos os conceitos (conceitos eleitos) a compor a OE, suas propriedades serão eleitas obedecendo a outras heurísticas, também detalhadas na seção 4.2.3. De forma resumida pode-se dizer que:

- para cada conceito eleito, todas as suas propriedades (atributos) serão eleitas e representadas na OE; e
- para cada conceito eleito, todas as suas propriedades (relações) para as quais há conceitos eleitos correspondentes, serão representadas na OE.

A ontologia gerada é armazenada no BD ONTOLOGIAS e reúne todos os conceitos e propriedades (atributos e relações) eleitos. Esta ontologia proporcionará subsídios iniciais e mais facilidade na concepção de uma ontologia final para a organização. Além disso, oferece uma visão global imediata para os usuários nos diversos pontos da Organização.

4.1.5 MÓDULO DE NAVEGAÇÃO-INSTANCIACÃO

A partir da solicitação do usuário, este módulo apresenta as instâncias de duas formas:

- *global* - tendo como referência a OE e utilizando os mapeamentos entre esta e as demais ontologias da organização para alcançar as instâncias dos demais pontos da organização.
- *local* - tendo como referência a ontologia local, e utilizando os mapeamentos realizados pelo ponto (entre os pontos) para alcançar instâncias dos outros pontos, e o mapeamento entre o ponto e a OE, e desta para os demais pontos, para alcançar instâncias de pontos ainda não mapeados localmente.

Para ilustrar o funcionamento deste módulo, serão consideradas quatro ontologias definidas por dois conceitos (sendo cada conceito representado por c_{jX} , onde j é a

identificação do conceito e X é a ontologia a qual pertence) e suas respectivas instâncias (representadas por i_{jXz} , onde z é a identificação da instância, j a identificação do conceito e X é a ontologia a qual pertence o conceito). Os mapeamentos entre ontologias serão representados por M_{XY} (conforme definição básica 7, seção 4.21) onde X é a ontologia fonte e Y a ontologia alvo (isto é, os conceitos das ontologias X são mapeados para os conceitos da ontologia Y) a saber: M_{12} (mapeamento entre ontologias O_1 e O_2), M_{24} , e M_{34} , e uma ontologia global OE gerada composta por dois conceitos, conforme apresentado na FIG. 4.5.

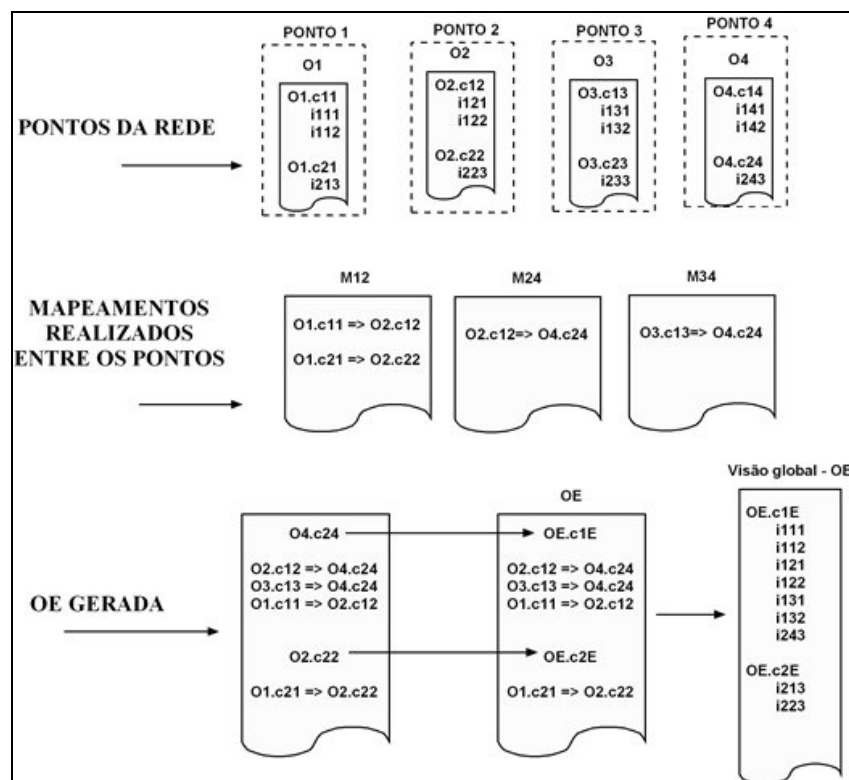


FIG. 4.5 Exemplo utilizado para demonstração do módulo de Navegação-Instanciação

Com base nestas ontologias e mapeamentos, é possível definir quais instâncias e conceitos que serão apresentadas ao usuário de forma Local e Global.

No caso da visão global serão apresentadas as instâncias dos conceitos eleitos $O4.c24$ e $O2.c22$ que passam a ser os conceitos $OE.c1E$ e $OE.c2E$, respectivamente e as instâncias relativas aos conceitos mapeados a partir de $OE.c1E$ e $OE.c2E$. Na FIG. 6.5 é pode-se ver as instâncias que serão apresentadas a partir da OE.

As instâncias $i141$ e $i142$, por exemplo, não serão apresentadas em virtude do conceito $O4.c14$ não ter mapeamentos com outros conceitos que o torne passível de ser um conceito eleito ou que proporcione a eleição de outro.

No caso do usuário do ponto 1, a visão local será constituída pelas instâncias apresentadas com base na própria ontologia do ponto (O_1) e do mapeamento realizado com a ontologia O_2 do ponto 2, formando as instâncias: i_{111} , i_{112} , i_{213} , i_{121} , i_{122} e i_{223} . Além disso, através dos mapeamentos de O_1 com a O_E , é possível visualizar ainda i_{131} , i_{132} , e i_{243} . Na FIG. 4.6 tem-se a visão local dos conceitos e as instâncias do ponto 1.

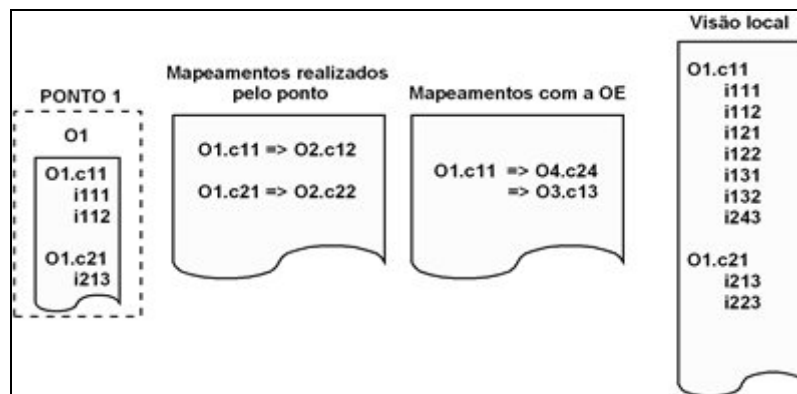


FIG. 4.6 Visão local do ponto 1

4.1.6 MÓDULO DE RASTREAMENTO

Este módulo realiza o monitoramento de algumas atividades dos demais módulos, mantendo arquivos de log. Os arquivos de log são arquivos-texto gerados pelo módulo de rastreamento têm por finalidade armazenar as seguintes informações:

- nome do usuário que iniciou a sessão;
- data de utilização do sistema;
- identificação dos mapeamentos realizados por um determinado usuário;
- situação da OE (disponível, indisponível, temporariamente indisponível, em-geração e instâncias-parciais);
- localização atual da OE (nome do ponto onde foi gerada); e
- pontos participantes da geração da OE atual;

O nome do usuário, data e as identificações dos mapeamentos realizados poderão ser utilizados pelo especialista do domínio, com a finalidade de consultar e esclarecer os mapeamentos realizados por determinado usuário.

A situação da OE é indicada de cinco maneiras: disponível, indisponível, temporariamente indisponível, em-geração e instâncias-parciais.. A situação *disponível* indica que a OE foi gerada e não será necessário que o módulo de gerência fique monitorando o BD MAP. A situação *indisponível* indica que a OE não foi gerada, cabendo ao módulo de Gerência monitorar o BD MAP. A situação *temporariamente indisponível* indica que o ponto que gerou a OE está temporariamente fora da rede P2P. Esta situação faz com que o módulo de Gerência apresente diretivas contingenciais ao usuário (buscar uma réplica da OE em outro ponto na rede, aguardar que o ponto que detém a OE reintegre a rede, ou gerar uma nova OE). A situação *instâncias-parciais* indica que existe algum ponto, que participou da geração da OE atual, temporariamente fora da rede P2P. Esta situação é necessária para que o módulo de navegação-instanciação avise ao usuário que algumas instâncias não poderão ser carregadas. A situação *em-geração* indica que a OE já está sendo gerada por algum ponto da rede, fazendo com que os demais aguardem a conclusão desse processo, e evitando com isso, processamento redundante.

4.1.7 BASES DE DADOS

O sistema possui três bases de dados, responsáveis por armazenar as ontologias, mapeamentos e instâncias. Estas bases utilizam diretórios do sistema de arquivos do sistema operacional e podem ser configuradas pelo módulo de gerência. As seguintes bases foram definidas:

- **BD MAP** - base de dados de mapeamentos realizados pelo ponto ou transferidos de outros pontos. Quando um ponto realiza um mapeamento, automaticamente, através do módulo de Comunicação e Gerência, este mapeamento é transferido o outro ponto envolvido;
- **BD ONTOLOGIAS** - base de dados onde são armazenadas a ontologia do ponto (local) e a OE; e
- **BD INSTÂNCIAS** - base de dados onde são armazenadas as instâncias referentes as ontologias e mapeamentos armazenados no *BD MAP* e no *BD ONTOLOGIAS*.

4.2 DEFINIÇÕES DE HEURÍSTICAS

O termo heurística é definido por um conjunto de regras e métodos que conduzem à descoberta, à invenção e à resolução de problemas. No contexto desta dissertação, as heurísticas têm como objetivo descobrir os conceitos e propriedades (atributos e relações) que irão compor a OE. A construção destas heurísticas possibilitou a formalização da proposta Ontologias Emergentes, viabilizando e facilitando, conseqüentemente, a implementação de um protótipo, denominado *OntoEmerge*.

Para que as heurísticas fossem formalizadas, foram inicialmente definidos os conceitos básicos, aqui denominados de definições básicas e definições de constantes, apresentadas nas seções 4.2.1 e 4.2.2.

As definições básicas retratam as diversas entidades envolvidas, como por exemplo: pontos de uma rede P2P, ontologias de um ponto, conjunto de todos os pontos de uma rede P2P, mapeamentos entre ontologias de dois pontos distintos, conjunto de mapeamentos, conjunto de mapeamentos armazenados em um determinado ponto, conceito de uma ontologia, conjunto de conceitos de uma ontologia, propriedade de uma ontologia, conjunto de propriedades de uma ontologia e outras. As definições de constantes declaram entidades, às quais serão atribuídos valores numéricos e alfabéticos, que serão configuradas em cada ponto, e que se manterão constantes para servir ao processo de geração da OE. Por fim, as heurísticas utilizadas são expressas através de equações, conforme descritas na seção 4.2.3.

4.2.1 DEFINIÇÕES BÁSICAS

Definição básica 1: Conjunto de pontos

Seja $P = \{p_1, p_2, \dots, p_x\}$, o conjunto de pontos em um ambiente P2P, onde x é o número de pontos deste ambiente.

Definição básica 2: Domínio de uma ontologia

Seja $D = \{d_1, d_2, \dots, d_k\}$ o conjunto de domínios de ontologias residentes em pontos de um conjunto P , onde k é o número de domínios.

Definição básica 3: Ontologia de um ponto

Seja $\Omega = \{O_1, O_2, \dots, O_n\}$, um conjunto de ontologias de um domínio k , residentes nos diversos pontos pertencentes a P . Consideramos que cada ponto p_x , onde $1 \leq x \leq n$, tem suas próprias ontologias, armazenadas localmente. Então tem-se que : $\{\forall p_x \in P, \text{ se } \exists O_i, O_j \wedge O_i, O_j \text{ são ontologias de um ponto } p_x, \text{ então } (O_i = O_j \wedge d_i = d_j) \vee (O_i \neq O_j \wedge d_i \neq d_j)\}$.

Em outras palavras, se um ponto possui mais de uma ontologia, então as ontologias são de domínios diferentes.

Definição básica 4: Conceito de uma ontologia

Seja c_{iX} , um conceito i da ontologia O_X , onde $1 \leq i \leq m$ e $1 \leq X \leq n$.

Definição básica 5: Conjunto de conceitos de uma ontologia O_X .

Seja $C_X = \{c_{iX}, \text{ onde } 1 \leq i \leq m \text{ e } 1 \leq X \leq n\}$, o conjunto de conceitos da ontologia O_X .

Definição básica 6: Mapeamento entre conceitos.

Seja a relação de mapeamento entre conceitos $map: C_X \rightarrow C_Y$. Diz-se que o conceito $c_{iX} \in C_X$, onde $1 \leq X \leq n$ é mapeado para o conceito $c_{jY} \in C_Y$, onde $1 \leq Y \leq n$, se $map(c_{iX}) = c_{jY}$.

A expressão m_{iXjY} será usada neste trabalho, por simplificação, para representar $map(c_{iX}) = c_{jY}$, sendo assim $m_{iXjY} = map(c_{iX}) = c_{jY}$.

Definição básica 7: Mapeamento entre ontologias.

Seja a relação de mapeamentos entre ontologias $Map: \Omega \rightarrow \Omega$. Diz-se que uma ontologia O_X , onde $1 \leq X \leq n$, está mapeada para outra ontologia O_Y , onde $1 \leq Y \leq n$, se $Map(O_X) = O_Y$.

A expressão M_{XY} será usada neste trabalho, por simplificação, para representar $Map(O_X) = O_Y$, sendo assim $M_{XY} \equiv Map(O_X) = O_Y$. M_{XY} pode ser definido ainda pelo par ordenado (O_X, O_Y) .

O mapeamento entre duas ontologias pode ser definido também, em termos dos mapeamentos entre os conceitos destas ontologias O_X e O_Y . Assim, podemos redefinir M_{XY} da seguinte forma:

Seja M_{XY} o conjunto de todos os mapeamentos entre os conceitos das ontologias O_X e O_Y , sendo que $M_{XY} \neq M_{YX}$, onde $1 \leq X \leq n$ e $1 \leq Y \leq n$. Assim sendo, tem-se que:

$$M_{XY} = \{ m_{iXjY}, \text{ onde } 1 \leq i \leq |C_X| \text{ e } 1 \leq j \leq |C_Y| / c_{iX} \in C_X \wedge c_{jY} \in C_Y \wedge map(c_{iX}) = c_{jY} \}$$

$$M_{XY} \neq M_{YX}$$

Definição básica 8: Pontuação de um mapeamento entre conceitos.

Seja a função de pontuação $\pi: M_{XY} \rightarrow R$, sendo R o conjunto dos números reais. Diz-se que um mapeamento m_{iXjY} tem pontuação r , se $\pi(m_{iXjY}) = r$. A expressão π_{iXjY} será usada neste trabalho, por simplificação, para representar $\pi(m_{iXjY})$, sendo assim: $\pi_{iXjY} \equiv \pi(m_{iXjY})$.

Definição básica 9: Pontuação de um conjunto de mapeamentos de conceitos realizados entre ontologias.

Seja a função $\Pi: \{M_{XY}: \forall O_X, O_Y \in \Omega\} \rightarrow R$.

Diz-se que a pontuação de um mapeamento de ontologia é a média das pontuações dos mapeamentos $m_{iXjY} \in M_{XY}$.

$$\Pi(M_{XY}) = \frac{\sum \pi_{iXjY}}{|M_{XY}|}, \text{ onde } 1 \leq i \leq |C_X| \text{ e } 1 \leq j \leq |C_Y|.$$

Definição básica 10: Mapeamento Inverso.

Diz-se que um mapeamento inverso é o mapeamento m'_{jYiX} que pode ser obtido a partir do mapeamento m_{iXjY} desde que $\pi(m_{iXjY}) = I$.

Por conseqüência, pode-se dizer também que um conjunto de mapeamentos M'_{YX} pode ser obtido a partir do M_{XY} , se $\forall m_{iXjY}$ de M_{XY} , $\pi_{iXjY} = I$.

Definição básica 11: Conjunto de Mapeamentos realizados entre as ontologias.

Seja M o conjunto de todos os mapeamentos realizados entre as ontologias $O_X \in \Omega$, onde $1 \leq X \leq n$. Tem-se então que: $M = \{M_{XY} / X \neq Y \wedge (O_X, O_Y \in \Omega)\}$

Definição básica 12: Conjunto de Mapeamentos envolvendo uma ontologia O_X .

Seja M_X o conjunto de todos os mapeamentos envolvendo uma ontologia O_X .

Tem-se então que: $M_X = \{M_{XY} / M_{XY} \in M, O_i, O_j \in \Omega \wedge (i = X \vee j = Y, X \neq Y)\}$.

Seja MT_{YX} o conjunto de mapeamentos de ontologia realizados para a ontologia O_X .

Tem-se então que: $MT_{YX} = \{MT_{YX} / MT_{YX} \in M, O_i, O_j \in \Omega \wedge (i = X \wedge j = Y)\}$.

Seja MF_{XY} o conjunto de mapeamentos de ontologia realizados a partir da ontologia O_X .

Assim, tem-se que: $MT_{YX} \cup MF_{XY} = M_X$

Definição básica 13: Conjunto de Mapeamentos realizados a partir de um conceito.

Seja MFC_{iX} o conjunto de mapeamentos (m_{iXjY}) realizados a partir do conceito c_{iX} da ontologia O_X .

Tem-se então que: $MFC_{iX} = \{M_{XY} / m_{iXjY}, \text{ onde } 1 \leq Y \leq n \text{ e } 1 \leq j \leq m\}$

Definição básica 14: Conjunto de Mapeamentos realizados para um conceito.

Seja MTC_{iX} o conjunto de mapeamentos (m_{jYiX}) realizados para um conceito c_{iX} da ontologia O_X .

Tem-se então que: $MTC_{iX} = \{ m_{jYiX}, \text{ onde } 1 \leq Y \leq n \text{ e } 1 \leq j \leq m \}$

Definição básica 14a: Caminho de mapeamento.

Diz-se que existe um caminho de mapeamento entre dois conceitos, quando existirem mapeamento entre ambos ou mapeamentos entre estes e outros conceitos de modo que possam ser alcançados.

Por exemplo, seja m_{iXjY} , m_{jYzV} , pode-se dizer que existe caminho de mapeamento entre os conceitos: c_{iX} e c_{zV} .

Definição básica 15: Conceito candidato

Um conceito candidato é um conceito que participou de um ou mais mapeamentos com outros conceitos e que pode ser então eleito para compor a OE. Ou seja, o conjunto de conceitos candidatos pode ser definido por

$$CC = \{c_{iX} \in C_X, O_X \in \Omega, \text{ onde } 1 \leq X \leq n \text{ e } 1 \leq i \leq m, \exists m_{iXjY} \vee m_{jYiX}\}.$$

Definição básica 16: Conceitos Afins

Conceitos Afins representam um conjunto de conceitos candidatos que têm algum vínculo de afinidade. Esta afinidade é definida levando-se em consideração os conceitos das diversas ontologias e seus mapeamentos. Pode-se dizer que dois conceitos são ditos conceitos afins quando existe um mapeamento entre eles, ou algum caminho de mapeamento (definição básica 14a) entre eles.

Na FIG. 4.7 existem vários mapeamentos, por exemplo, m_{iXjY} , m_{jYiX} e m_{iXzZ} entre os conceitos. Nesta figura, nota-se que c_{iX} , c_{jY} e c_{zZ} são conceitos afins, assim como c_{2X} , c_{3Z} e c_{3Y} e ainda, c_{5X} , c_{4Z} e c_{4Y} .

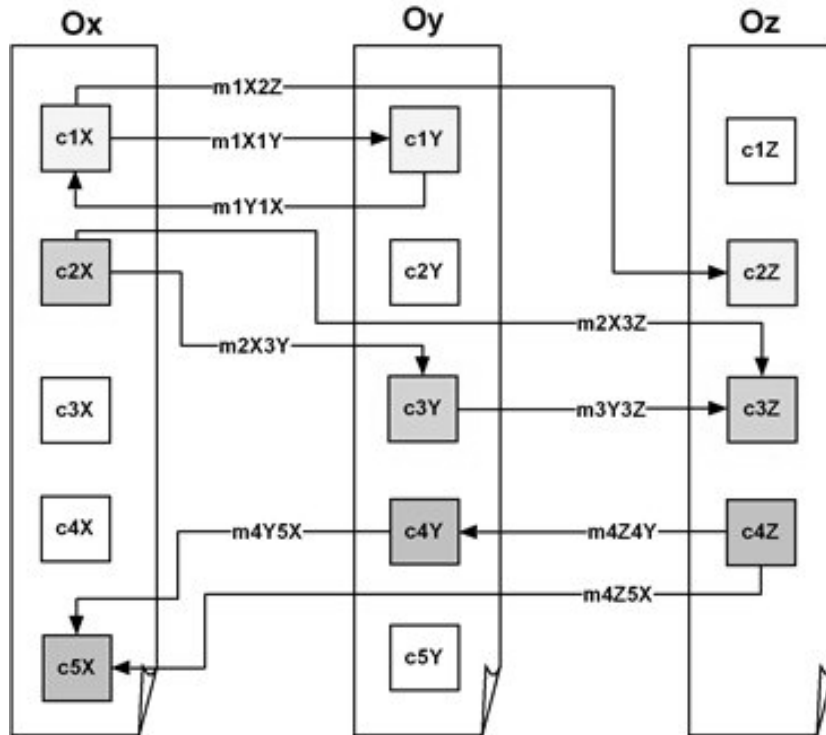


FIG. 4.7 Exemplo de mapeamentos entre conceitos de Ontologia, que formam um grafo conexo

Utilizando-se a teoria de grafos para expressar com maior clareza o conceito de afinidade, considera-se os mapeamentos como as arestas de um grafo e os conceitos como vértices. Definem-se os conjuntos de Conceitos Afins como se segue: seja um grafo $G(V,E)$, onde $E = \{(c_{iX}, c_{jY}) / \exists m_{iXjY} \vee m_{jYiX}\}$ e $V = \{c_{iX} \in C_X, \forall C_X / O_X \in \Omega\}$. Os conceitos dos subgrafos de G que são conexos formam os diversos conjuntos de conceitos afins, CA_i . Nota-se que na FIG. 4.5, formaram-se 3 grafos conexos: $CA_1 = \{c_{1X}, c_{1Y}, c_{2Z}\}$, $CA_2 = \{c_{2X}, c_{3Y}, c_{3Z}\}$ e $CA_3 = \{c_{5X}, c_{4Y}, c_{4Z}\}$.

Definição básica 17: Conjunto de conceitos eleitos

Seja $CE = \{c_{1Y}, c_{3Z}, c_{1G}, c_{3X}, c_{4X}, \dots, c_{mn}\}$ o conjunto de conceitos, isto é, selecionados através das heurísticas, definidas na seção 4.2.3, para composição da OE .

Definição básica 18: Classificação das ontologias - g_X

Cada O_X em Ω é desenvolvida por um grupo da organização. Considerando que cada grupo tem um nível diferente de conhecimento sobre o domínio da Ontologia criada, atribuímos uma classificação a estes grupos, conforme o nível de conhecimento. Isto é, quanto maior o conhecimento sobre o domínio, maior a classificação atribuída. Assim, dependendo da organização é possível, por exemplo, definir as seguintes classes:

- Especialistas do domínio (ED) - Grupo de usuários que detêm a grande parte do conhecimento sobre determinado domínio;
- Especialistas do subdomínio (ES) - Grupo de usuários que detêm o conhecimento de parte de um domínio; e
- Especialistas da Ciência da Informação (ECI) – grupo de usuários que embora não sejam especialistas no domínio, estão trabalhando na definição de ontologias de domínio junto a especialistas de domínio.

Considerando estas classes, pode-se dizer que ECI tem maior nível de conhecimento que ED, e que ED tem maior nível de conhecimento que ES. Logo, tem-se que $ECI > ED > ES$.

Diz-se então que $g(O_X)$, ou simplesmente g_X , é a classificação atribuída a O_X , considerando o nível de conhecimento do grupo que desenvolveu O_X sobre o domínio em questão.

Definição básica 19: Ranking de um conceito

O ranking de um conceito é calculado com base na abordagem denominada de *PageRank* (BRIN & PAGE, 1998) utilizada pela ferramenta de busca Google. O *PageRank* tem como principal característica a utilização da estrutura de *links* de páginas *web*. A ferramenta Google interpreta, por exemplo, um *link* da página A para a página B, como um voto da página A para a página B. Os votos dados pelas páginas que são por si próprias, "importantes", têm maior peso e ajudam a tornar outras páginas "importantes". As páginas mais relevantes serão aquelas que obtiverem o maior *PageRank*.

Considerando-se que uma determinada página A tenha $T1$ a Tn páginas que apontam para a mesma, que o parâmetro d seja o *fator de amortecimento*, o qual poderá ter valor entre 0 e 1; além disso, seja $C(Tn)$ definido como o número de links que saem da página n . Diante do exposto acima, o *PageRank* de A será dado pela equação:

$$PR(A) = (1-d) + d (PR(T1)|C(T1) + \dots + PR(Tn)|C(Tn)) \text{ (eq. 1)}$$

O parâmetro d permite a personalização da página e evita que alguns usuários, deliberadamente, enganem o sistema e consigam valor de ranking maior.

No contexto deste trabalho os conceitos das ontologias foram utilizados no lugar das páginas web. Se o conceito A é mapeado para o conceito B, então B tem um voto de A. O conceito que tiver o maior *PageRank* será eleito para compor a OE. Adaptando-se a equação (eq. 1) para o contexto de mapeamentos de conceitos, considera-se para um dado conceito c_{iX} , o conjunto de conceitos c_{jY} para os quais existe mapeamento, isto é, $map(c_{iX}) = c_{jY}$.

Tem-se então, a seguinte equação resultante:

$$\rho(c_{iX}) = (1-d) + d * (\sum \rho(c_{jY}) \mid \mid MFC_{jY} \mid \mid) \text{ (eq. 2)}$$

Onde:

$$1 \leq X \leq n \text{ e } 1 \leq i \leq m;$$

$$1 \leq Y \leq n \text{ e } 1 \leq j \leq m;$$

$\mid MFC_{jY} \mid$ - número de mapeamentos realizados a partir do conceito c_{jY} (definição 13); e

d - é o fator de amortecimento, que no contexto deste trabalho, é baseado na classificação dada à ontologia (definição básica 18), a qual pertence o conceito c_{iX} .

No caso das páginas Web, para o cálculo do *PageRank*, foram considerados, basicamente, o fator de amortecimento d e a quantidade de links que saem de uma página. Foi observado que não são atribuídas pontuação para cada *link*, isto é, quando uma página aponta para outra, não há pesos diferentes para os votos computados. Entretanto, no presente trabalho, quanto um conceito é mapeado para outro, este mapeamento tem uma pontuação associada, que foi considerada relevante para o cálculo do *ranking* de um conceito. Assim, para levar em conta a pontuação dos mapeamentos, em (eq. 2) tem-se a seguinte expressão resultante (eq. 3):

$$\rho(c_{iX}) = (1-d) + d * (\sum \pi_{jYiX} * \rho(c_{jY}) \mid \mid MFC_{jY} \mid \mid), \text{ (eq. 3)}$$

Onde:

$$1 \leq X \leq n \text{ e } 1 \leq i \leq m;$$

$$1 \leq Y \leq n \text{ e } 1 \leq j \leq m;$$

$|MFC_{jY}|$ - número de mapeamentos realizados a partir do conceito c_{jY} (definição 13);
 d - é o fator de amortecimento, que no contexto deste trabalho, está baseado na
classificação dada à ontologia a qual pertence o conceito c_{iX} ; e
 π_{jYiX} - pontuação do mapeamento de um conceito c_{jY} para c_{iX} .

Definição básica 20: Atributo de um conceito c_{iX}

Atributos são propriedades que expressam características ou aspectos que um dado conceito pode assumir.

Considerando o conceito c_{iX} , diz-se que pa_{iXz} representa o atributo z de c_{iX} .

Definição básica 21: Relacionamento entre conceitos c_{iX} e c_{jX} .

Propriedades expressam relacionamentos entre conceitos de uma mesma ontologia.

Sejam c_{iX} e c_{jX} dois conceitos de O_X . Diz-se que $pr_z(c_{iX}, c_{jX})$, ou simplesmente pr_{ijXz} representa o relacionamento z de c_{iX} com c_{jX} .

Nas linguagens usadas para a representação de ontologias, a exemplo de RDF e OWL, o conceito de *domínio* é usado para restringir o conjunto de valores que possuem uma determinada propriedade e o conceito de *range* é usado para restringir o conjunto de valores válidos para se atribuir a uma determinada propriedade (RDFS).

Assim, sejam as funções: $D:pr_z \rightarrow c_{iX}$ e $R:pr_z \rightarrow c_{jX}$ que representam, respectivamente, o domínio e o *range* do relacionamento z . Então, tem-se que: $D(pr_z) = c_{iX}$ e $R(pr_z) = c_{jX}$.

Definição básica 22: Conjunto de propriedades de um conceito c_{iX}

Seja $Pr_{iXj} = \{pr_{ijXz}, pa_{iXa}, pa_{iXb} \dots\}$ o conjunto de propriedades j de um conceito c_{iX} que é constituído de seus atributos e seus relacionamentos.

Definição básica 23: Conjunto de propriedades eleitas

Seja $PE = \{ pr_{ijXz}, pa_{iZy}, pa_{iXz}, \dots \}$ o conjunto de propriedades eleitas, escolhidas para compor a OE, conforme as heurísticas definidas na seção 4.2.3.

4.2.2 DEFINIÇÕES DE CONSTANTES

Definição constante 1:

Seja k_I a quantidade mínima de mapeamentos realizados entre ontologias. Pode-se definir como *default* que uma quantidade mínima de mapeamentos poderia ser o número de ontologias envolvidos menos 1, considerando que entre cada par de ontologias existiria pelo menos um mapeamento. Então, é possível assumir como default que $k_I = |\Omega| - 1$.

Definição constante 2: Mapeamento de alta pontuação

Existem várias referências bibliográficas CMS (KALFOGLOU & HU, 2005), AUTOMS (KOTIS et al, 2006) e S-MATCH (GIUNCHIGLIA et al, 2004) que tratam de similaridades e relacionamentos semânticos entre conceitos. Diante da análise realizada em (LACTHIM et al, 2008) verificamos a existência de ferramentas que utilizam algoritmos de cálculo de similaridade entre conceitos, visando propor casamentos.

O valor que define se dois conceitos são similares depende da ferramenta de cálculo de similaridade utilizada. Em (KALFOGLOU & HU, 2005), por exemplo, este valor é limitado ao intervalo (0,1). Quanto mais próximo de 1, mais similares são os conceitos. Os autores chegam a sugerir que dois conceitos podem ser considerados similares se o cálculo da similaridade pontuar acima de 0,80 CMS (KALFOGLOU & HU, 2005) ou 0,75 OMEN (MITRA et al., 2004).

Diante do exposto anteriormente, é possível concluir que a pontuação do mapeamento poderia ser determinada manualmente ou através de mecanismos automáticos. Assim, no escopo desta dissertação, mapeamentos de alta pontuação serão definidos de acordo com a definição a seguir:

Seja um mapeamento m_{jYIX} tal que a sua pontuação seja π_{jYIX} . A constante k_2 define o limite mínimo para que um mapeamento seja considerado de alta pontuação. Então pode-se dizer que se $\pi_{jYIX} \geq k_2$ tem-se um mapeamento de alta pontuação.

4.2.3 HEURÍSTICAS PARA COMPOSIÇÃO DA OE

A heurística 1 quando satisfeita, dá início ao processo de geração da OE. Para gerar a OE é necessário que existam mapeamentos que envolvam todos os pontos.

A heurística 2 visa definir os conceitos que compõem a OE. Esta verifica qual o conceito entre os Conceitos Afins será eleito para compor a OE, fazendo isso para todos os conjuntos de conceitos afins. Por fim, as heurísticas 3 e 4 verificam quais as propriedades dos conceitos eleitos pela heurística 2 devem ser representadas na OE.

Definição heurística 1: Verificação do momento inicial para geração da OE

Esta heurística é usada para verificar se os mapeamentos englobam todos os pontos. A idéia é verificar se a organização alcançou um nível mínimo de troca de informações, a ponto de possibilitar a composição da OE a partir destas trocas. Isto é, o objetivo é garantir que partindo de um ponto qualquer, através dos mapeamentos (desconsiderando a direção dos mesmos), se chegaria a qualquer outro ponto da organização.

Utilizou-se aqui, a teoria de grafos para expressar com maior clareza esta heurística. No exemplo da FIG. 4.8, cada ponto será considerado um vértice do grafo e os mapeamentos existentes entre os pontos as arestas (desconsiderando-se a direção). Se o resultado desta composição formar um grafo conexo, é sinal de que tem-se todos os pontos envolvidos nos mapeamentos. Pode-se observar quatro pontos na FIG. 4.8. Para que os mesmos formem um grafo conexo, a quantidade mínima de mapeamentos requeridos será de 3, isto é: $|M| = k_I$, sendo $k_I = |\Omega| - 1$.

Generalizando, considere o grafo $G(V,E)$, onde $V = \Omega = \{O_1, O_2, \dots, O_n\}$ e $E = \{M_{ij} / i \neq j \wedge (O_i, O_j \in \Omega)\}$. Se G é conexo, então tem-se informações suficientes para analisar os mapeamentos entre conceitos necessários à geração da OE em um determinado ponto p_X .

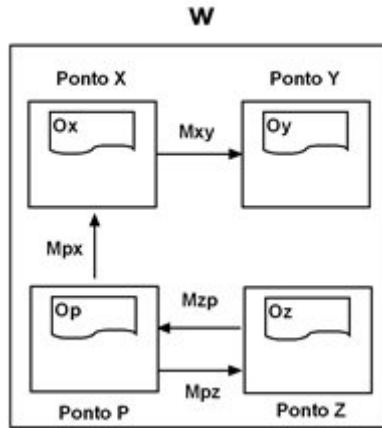


FIG. 4.8 Exemplo de mapeamentos entre ontologias envolvendo quatro pontos

Definição heurística 2: Eleição de Conceitos para a OE

Considera os *rankings* calculados (definição básica 19) para cada conceito candidato dentro de um conjunto de conceitos afins (definição básica 16). Esta heurística visa eleger o conceito de maior *ranking*, que representará na OE os conceitos afins pertencentes a este conjunto.

Assim, tem-se que: Dado um $c_{iX} \in CA_k$, se $\forall c_{jY} \in CA_k / c_{jY} \neq c_{iX} \wedge \rho(c_{iX}) > \rho(c_{jY})$, então $c_{iX} \in CE$ sse $\forall c_{jY} \in CA_k, c_{jY} \neq c_{iX} \wedge \rho(c_{iX}) > \rho(c_{jY})$,

Definição heurística 3: Eleição de Atributos para os conceitos da OE

Esta heurística visa verificar quais atributos de um determinado conceito eleito, serão representados na OE.

Para todo $c_{iE} \in CE$, qualquer que seja o atributo de c_{iX} , tal que $c_{iE} = c_{iX}$, este atributo será também de c_{iE} . Assim, tem-se que: $\forall c_{iE} \in CE / (c_{iE} = c_{iX})$ se $(\exists pa_{iXz} \in Pr_{iXj})$ então $(pa_{iXz} \in Pr_{iEj}) \wedge (pa_{iXz} \in PE)$

Definição heurística 4: Eleição de Propriedades para os conceitos da OE

Esta heurística tem por propósito definir os relacionamentos de um conceito eleito. Dado um conceito eleito c_{iE} equivalente a c_{iX} , para toda propriedade pr_z cujo domínio seja c_{iX} , isto é, $(D(pr_z) = c_{iX})$, se $range R(pr_z) = c_{jX}$ é equivalente a um conceito eleito c_{jE} , ou faz parte de um CA_k para o qual há um conceito eleito associado c_{jE} , então a propriedade pr_z será representada na OE da seguinte forma: $D(pr_z) = c_{iE}$ e $R(pr_z) = c_{jE}$.

Assim, tem-se que:

seja $(c_{iE}, c_{jE} \in CE) / (c_{iE} = c_{iX}) \wedge (\exists pr_z \in Pr_{iX} / D(pr_z) = c_{iX} \wedge R(pr_z) = c_{jX})$

se $(c_{jE} = c_{jX}) \vee (c_{jE} = c_{jY} / c_{jX}, c_{jY} \in CA_k)$

então $(pr_z \in Pr_{iEj} / D(pr_z) = c_{iE} \wedge R(pr_z) = c_{jE}) \wedge (pr_z \in PE)$

4.3 CONSIDERAÇÕES FINAIS

Este capítulo teve por objetivo apresentar a arquitetura da abordagem de Ontologias Emergentes, que consiste de cinco módulos e três bases de dados, bem como descrever o detalhamento das funcionalidades, complementando-as com a definição das heurísticas necessárias no desenvolvimento do sistema.

O uso de definições em linguagem formal visando representar as entidades envolvidas na geração da OE foi necessário e facilitou a criação das heurísticas. Algumas definições básicas a exemplo de: conjunto de ontologias, conjunto de conceitos, conjunto de mapeamentos, conceitos afins e outros, assim como a especificação de heurísticas de mapeamento, até onde se investigou, não foi encontrada na literatura. Dessa forma, considera-se esta etapa como uma contribuição importante desta dissertação para a área, pois facilitará a especificação de outras propostas e mecanismos de interoperabilidade de ontologias.

Por meio do sistema especificado foi possível obter uma estrutura inicial para gerar uma visão global da organização baseado na troca de informações, retratada pelos mapeamentos e adaptando a idéia do *PageRank* descrita em (BRIN & PAGE, 1998). A escolha do *PageRank* para a obtenção dos conceitos eleitos, que compõe a heurística 2, foi baseada em uma analogia feita entre o ato de fazer referência a uma página na Web e o ato de fazer um

mapeamento para um conceito de uma ontologia. No entanto, acredita-se que esta heurística poderia ser melhor refinada, com base em uma análise qualitativa mais aprofundada.

Vale destacar ainda que a OE gerada não pode ser considerada com uma visão geral definitiva da organização, mas sim como uma base para a construção de uma ontologia organizacional inicial, que possa auxiliar uma equipe especializada.

A teoria de grafos foi utilizada principalmente para expressar com maior clareza algumas heurísticas e definições básicas. Por exemplo, na definição básica 16 apresentada na seção 4.2.1, a idéia de colocar todos os conceitos como vértices de um grafo e os mapeamentos entre os conceitos como arestas, proporcionou um melhor entendimento de como seriam separados os conceitos.

O capítulo seguinte apresenta o desenvolvimento do protótipo *OntoEmerge*, gerado com base na especificação da abordagem de Ontologias Emergentes, com o detalhamento de implementação dos módulos e bases de dados. A implementação foi realizada a partir da utilização de diversas ferramentas, algumas das quais já extensamente aplicadas em outros sistemas da comunidade acadêmica, e tendo como fundamento a aplicação de todas as heurísticas criadas ao longo deste capítulo. O enfoque dado para a programação se restringiu essencialmente aos módulos de Geração OE e de Gerência (FIG. 4.1) em virtude algumas funcionalidades terem sido já resolvidas por outras ferramentas como o MAFRA (MAEDCHE et al., 2002) e (NUNO SILVA, 2004) e o COPPEER (MIRANDA, 2006).

5 PROTÓTIPO DO SISTEMA

Este capítulo tem como objetivo apresentar a implementação do protótipo do sistema *OntoEmerge*, tendo como base a arquitetura especificada anteriormente no capítulo 4 e os estudos das abordagens descritas nos capítulos 2 e 3.

Inicialmente, será mostrada uma visão geral do protótipo do sistema *OntoEmerge* seguida posteriormente de uma descrição mais detalhada de sua implementação.

5.1 VISÃO GERAL DO PROTÓTIPO

A implementação das funcionalidades previstas na especificação do sistema *OntoEmerge* priorizou aquelas necessárias para a geração de uma visão global a partir dos mapeamentos disponíveis em um ponto. Assim, o protótipo desenvolvido viabilizou a aplicação de todas heurísticas especificadas.

A arquitetura do sistema apresentada na FIG. 5.1 fornece uma visão geral dos seus componentes, embora somente alguns deles tenham sido implementados. O módulo de Geração da OE foi desenvolvido, enquanto que o módulo de Gerência foi apenas parcialmente implementado. Os módulos de Rastreamento e Navegação-Instanciação também não foram implementados. As funcionalidades previstas para o módulo de Mapeamento P2P foram resolvidas pela ferramenta de mapeamento MAFRA (NUNO SILVA, 2004) e (MAEDCHE et al., 2002). Já o módulo de Comunicação e Gerência P2P foi desenvolvido através da ferramenta COPPEER (MIRANDA, 2006), sendo que sua funcionalidade básica é a de transferência de arquivos entre pontos. As bases de dados BD ONTOLOGIAS, BD MAP e BD INSTÂNCIAS são diretórios do sistema operacional, e são usadas para armazenar os arquivos que descrevem as ontologias, os mapeamentos entre elas e suas instâncias respectivamente.

A FIG. 5.1 re-apresenta a arquitetura conceitual do *OntoEmerge* (já apresentada no capítulo 4), que é idêntica para todos os pontos disponíveis na rede, e evidencia o que está funcional no protótipo, e o que não foi implementado.

A plataforma de desenvolvimento utilizada foi a proposta pela comunidade de *software* livre, denominada Eclipse (ECLIPSE) versão 3.3.1.1, e a linguagem utilizada Java (JDK 6.0).

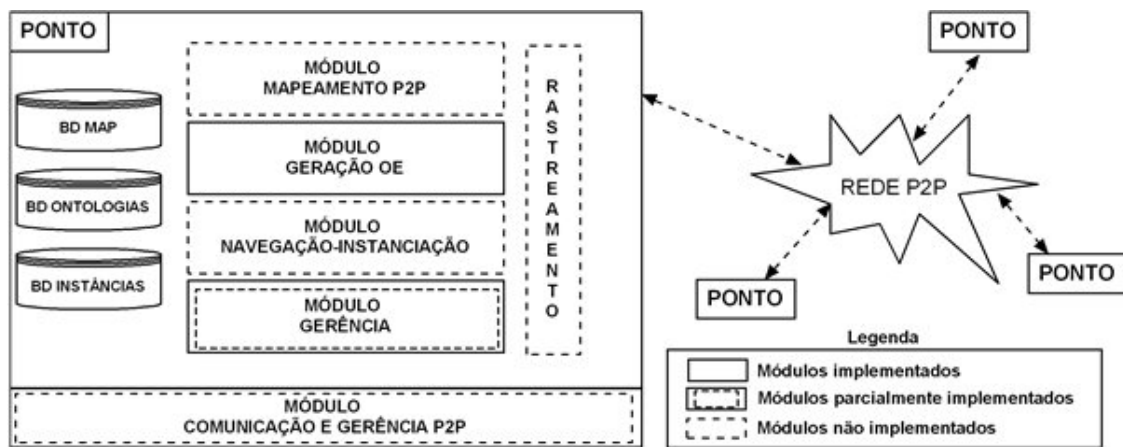


FIG. 5.1 Arquitetura conceitual do OntoEmerge

5.2 IMPLEMENTAÇÃO A PARTIR DOS MÓDULOS DA ARQUITETURA (DIAGRAMAS)

Para a descrição dos módulos desta arquitetura, tornou-se necessária a criação de algumas ontologias para melhor exemplificar suas funcionalidades. Assim, foram criadas as ontologias CPAD, LEGISMAR, SIGDEM e PESSOAL cuja descrição detalhada encontra-se no capítulo 6.

5.2.1 MÓDULO DE MAPEAMENTO PONTO-A-PONTO

Este módulo é responsável pela realização de mapeamentos entre ontologias. Pressupõe-se que cada ponto utilize a ferramenta de mapeamento MAFRA, e que para cada mapeamento, seja gerado um arquivo de mapeamento no formato SBO proposto por (NUNO SILVA, 2004) e (MAEDCHE et al., 2002), na linguagem RDF. Este arquivo em RDF contém as informações dos mapeamentos realizados.

Considere, por exemplo, duas ontologias denominadas de SIGDEM e CPAD, constituídas de classes representadas em RDF conforme apresentado na FIG. 5.2 e FIG. 5.3, respectivamente. Os mapeamentos realizados entre estas ontologias são representados pelo MAFRA em um arquivo RDF, conforme mostrado na FIG. 5.4.

```

...
1 <rdfs:Class rdf:about="&SIGDEM;EXPEDIENTE"
2   rdfs:label="EXPEDIENTE">
3   <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
4 </rdfs:Class>
...

```

FIG. 5.2 Representação da Classe EXPEDIENTE da ontologia SIGDEM em RDF

```

...
1 <rdfs:Class rdf:about="&CPAD;DOCUMENTOS"
2   rdfs:label="DOCUMENTOS">
3   <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
4 </rdfs:Class>
...

```

FIG. 5.3 Representação da Classe DOCUMENTOS da ontologia CPAD em RDF

```

...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file:/D:/MAPS/SIGDEM.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file:/D:/MAPS/CPAD.rdfs"/>
4 </a:Mapping>
5 <a:ConceptBridge rdf:ID="i-1193416327078-710427541">
6   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/SIGDEM#EXPEDIENTE"/>
7   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
8 </a:ConceptBridge>
...

```

FIG. 5.4 Informação de mapeamento da ontologia SIGDEM para CPAD

Nessa FIG. 5.4 tem-se a representação de um mapeamento (*Mapping* - linha 1) entre as duas ontologias, denominadas fonte (*relatesSourceOntology* - linha 2) e alvo (*relatesTargetOntology* - linha 3), indicando que a ontologia fonte foi mapeada para ontologia alvo. Assim, é possível observar que cada ontologia envolvida pertence a um determinado ponto.

Neste arquivo de mapeamento são registradas também informações sobre o mapeamento entre conceitos. O formato SBO representa o que se chama de ponte entre conceitos (*ConceptBridge*), onde cada ponte gerada refere-se a dois conceitos que foram mapeados. Na FIG. 5.4 tem-se a representação do mapeamento entre conceitos, na qual o conceito fonte (*relatesSourceEntity* - linha 6) *EXPEDIENTE* da ontologia SIGDEM foi mapeado para o conceito alvo (*relatesTargetEntity* - linha 7) *DOCUMENTOS* da ontologia CPD.

A SBO não contempla a informação da medida de similaridade dos mapeamentos realizados entre os conceitos de cada ontologia, sendo assim foi necessária a colocação de um elemento, aqui chamado de *PontuacaoMap*, para registrar estes dados, sendo assim a SBO passou a ser denominada de *SBO estendido*.

A FIG. 5.5 apresenta o formato *SBO estendido* com a informação da medida de similaridade no elemento *PontuacaoMap* (linha 8).

```
...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file://D:/MAPS/LEGISMAR.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file://D:/MAPS/SIGDEM.rdfs"/>
4 </a:Mapping>
5 <a:ConceptBridge rdf:ID="i-1193416974000-1073996323">
6   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#DOCUMENTOS"/>
7   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/SIGDEM#EXPEDIENTE"/>
8   <ime:PontuacaoMap>1.0^^http://www.w3.org/2001/XMLSchema#float</ime:PontuacaoMap>
9 </a:ConceptBridge>
...
```

FIG. 5.5 Representação de mapeamentos entre ontologias - SBO estendido

As informações no arquivo de mapeamento no elemento *Mapping* será utilizado pelo módulo de Gerência e as informações nos elementos *PontuacaoMap* e *ConceptBridge* serão utilizadas pelo módulo de Geração da OE. Outras informações também são registradas neste arquivo como, por exemplo: os mapeamentos entre propriedades e os serviços de transformações, mas não serão utilizadas pelo protótipo proposto.

Os mapeamentos realizados (arquivos de mapeamento) pelo ponto ou recebidos de outros pontos são armazenados na base de dados BD MAP.

5.2.2 MÓDULO DE GERÊNCIA

O módulo de gerência utiliza as informações de mapeamentos representadas nos arquivos de mapeamento (arquivos SBO/RDF), armazenadas no BD MAP, e aplica a heurística 1, definida no capítulo 4, para verificar o momento de início de geração da OE.

Para aplicação da heurística 1, o módulo utiliza as informações constantes nos sub-elementos *relatesSourceOntology* e *relatesTargetOntology* do elemento *Mapping* (FIG. 5.4), onde estes sub-elementos passam a ser representados por vértices de um grafo e seus mapeamentos (desconsiderando a direção) passam a ser representados pelas arestas que

conectam os vértices. Todos os arquivos de mapeamento entre ontologias são analisados e por fim obtém-se um grafo que representa todos os mapeamentos já realizados. A heurística 1 é satisfeita quando este grafo é conexo. Caso isso ocorra, o módulo de Gerência sinaliza para o módulo de Geração da OE, para que este prossiga na aplicação das demais heurísticas.

Para a representação do grafo com os diversos mapeamentos de ontologias existentes foi utilizada uma estrutura de matriz de adjacência (simétrica). Esta matriz possui uma linha com os nomes das ontologias fonte e uma coluna com os nomes das ontologias alvo, e caso existam mapeamentos entre as ontologias, o valor de cada par (coluna, linha) será verdadeiro (*true*) e, em caso contrário, será falso (*false*).

As informações sobre mapeamentos entre ontologias são obtidas de todos os arquivos de mapeamento existentes no ponto, e a partir destas informações monta-se a matriz de adjacências, descrita na tabela 5.1

Ontologia Alvo ↓	Ontologia Fonte →	/D:/MAPS/CPAD.rdfs	/D:/MAPS/LEGISMAR.rdfs	/D:/MAPS/PESSOAL.rdfs	/D:/MAPS/SIGDEM.rdfs
/D:/MAPS/CPAD.rdfs		<i>False</i>	<i>True</i>	<i>True</i>	<i>False</i>
/D:/MAPS/LEGISMAR.rdfs		<i>True</i>	<i>False</i>	<i>False</i>	<i>True</i>
/D:/MAPS/PESSOAL.rdfs		<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>
/D:/MAPS/SIGDEM.rdfs		<i>False</i>	<i>True</i>	<i>False</i>	<i>False</i>

TAB. 5.1 Matriz de adjacência representativa dos mapeamentos existentes em um ponto

Diante dos mapeamentos entre ontologias representados através desta matriz, para verificação de um da condição imposta pela heurística 1, foi aplicada a busca em profundidade na matriz, visando auxiliar a obtenção das informações de conexidade do grafo. A FIG. 5.6 a seguir representa a matriz da tabela 5.1 na forma de um grafo. Neste exemplo nota-se que o grafo é conexo.

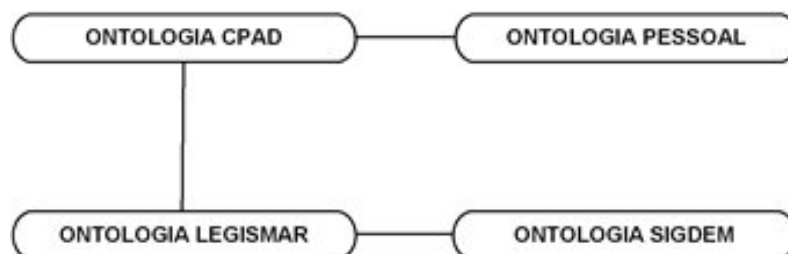


FIG. 5.6 Grafo dos mapeamentos realizados entre as ontologias

Para manipulação e seleção das informações de mapeamentos (*Mapping*) existentes nos arquivos no formato SBO/RDF foi utilizada a ferramenta Jena (JENA), que possui mecanismos de suporte à linguagem SPARQL (SPARQL).

A interface de configuração prevista na especificação não foi implementada. Assim, os parâmetros necessários para o funcionamento deste módulo foram colocados diretamente na programação, como constantes. A interface de configuração, quando implementada, produzirá um arquivo XML válido, de acordo com o esquema XML, conforme mostra a FIG. 5.7. Este arquivo armazena as informações de configuração utilizadas pelo módulo de Gerência, módulo de Geração da OE, módulo de Navegação-Instanciação e o módulo de Rastreamento.

No caso do módulo de Gerência, os elementos utilizados são: SITUACAO-OE, QTD-PONTOS, BD-MAP e TEMPO-MONITORAMENTO.

O elemento NOME-PONTO define o nome identificador do ponto, que será consultado pelo módulo de Rastreamento.

O elemento SITUACAO-OE define a situação da OE, sendo que o módulo responsável por modificar esta situação é o módulo de Gerência e Geração da OE. O módulo de Gerência realiza a consulta a este elemento para verificar se a OE está sendo gerada ou se já existe em outro ponto. Caso o elemento SITUACAO-OE esteja com uma instância igual a "disponível" (descrita na seção 4.1.6 do capítulo 4), o módulo de Gerência cessa o monitoramento do BD MAP.

O elemento QTD-PONTOS define a quantidade de pontos na rede que utilizam o sistema. O módulo de Gerência consulta então estas informações, com a finalidade de calcular a quantidade mínima de mapeamentos entre as ontologias dos pontos existentes. O módulo de Gerência e Comunicação P2P realiza a atualização deste valor no momento da entrada de um ponto na rede.

O elemento BD-INSTANCIAS define o diretório do sistema operacional onde serão armazenadas as instâncias dos mapeamentos e da OE. O módulo de Comunicação e Gerência P2P consulta esta informação, com o objetivo de verificar onde serão armazenadas as instâncias que, posteriormente, serão acessadas pelo módulo de Navegação-Instanciação.

O elemento BD-ONTOLOGIAS define o diretório do sistema operacional onde serão armazenadas as ontologias e a OE. O módulo de Geração da OE consulta esta informação para verificar onde a OE será armazenada, e o módulo de Comunicação e Gerência P2P consulta esta informação com o objetivo de verificar onde serão armazenadas a ontologia do ponto e as ontologias transferidas dos outros pontos.

O elemento BD-MAP define o diretório do sistema operacional onde serão armazenados os mapeamentos realizados e/ou recebidos pelo ponto. Os módulos de Gerência, Navegação-Instanciação e Comunicação e Gerência P2P consultam estas informações.

O elemento TEMPO-MONITORAMENTO define o intervalo de tempo em que o módulo de Gerência deve realizar a análise dos arquivos de mapeamentos armazenados no BD MAP, com a finalidade de verificar a viabilidade de geração da OE.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified"
  attributeFormDefault="unqualified">
- <xs:element name="NOME-PONTO">
  - <xs:annotation>
    <xs:documentation>Nome identificador do ponto</xs:documentation>
  </xs:annotation>
  </xs:element>
- <xs:element name="SITUACAO-OE">
  - <xs:annotation>
    <xs:documentation>Situacao que a OE se encontra no ponto - disponível, indisponível, temporariamente
    indisponível, em-geração ou instâncias-parciais</xs:documentation>
  </xs:annotation>
  </xs:element>
- <xs:element name="QTD-PONTOS">
  - <xs:annotation>
    <xs:documentation>Quantidade de pontos participantes da rede</xs:documentation>
  </xs:annotation>
  </xs:element>
- <xs:element name="BD-INSTANCIAS">
  - <xs:annotation>
    <xs:documentation>Diretório de do sistema operacional, onde serao armazenadas as
    instancias</xs:documentation>
  </xs:annotation>
  </xs:element>
- <xs:element name="BD-ONTOLOGIAS">
  - <xs:annotation>
    <xs:documentation>Diretório de do sistema operacional, onde serao armazenadas ontologias e a
    OE</xs:documentation>
  </xs:annotation>
  </xs:element>
- <xs:element name="BD-MAP">
  - <xs:annotation>
    <xs:documentation>Diretório de do sistema operacional, onde serao armazenados os mapeamentos
    realizados/recebidos</xs:documentation>
  </xs:annotation>
  </xs:element>
- <xs:element name="TEMPO-MONITORAMENTO">
  - <xs:annotation>
    <xs:documentation>Intervalo de tempo de monitoramento do BD MAP</xs:documentation>
  </xs:annotation>
  </xs:element>
</xs:schema>

```

FIG. 5.7 Arquivo produzido pela interface de configuração

5.2.3 MÓDULO DE GERAÇÃO DA OE

O módulo de geração da OE utiliza as informações do elemento *ConceptBridge*, representadas nos arquivos de mapeamento (veja um exemplo na FIG. 5.8), e aplica a heurística 2 para eleger os conceitos que serão representados na OE. As heurísticas 3 e 4 serão utilizadas também para eleger as propriedades dos Conceitos Eleitos (definição básica 17), igualmente representados na OE.

Devido a pouca dinamicidade dos mapeamentos entre as ontologias, a repetição de execuções da equação 3 (definição básica 19), por este módulo, influencia pouco no resultado, isto é, na escolha o conceito.

Após observações funcionais de iteração da equação 3 no *OntoEmerge*, nota-se que os conceitos escolhidos foram pouco alterados. Optou-se por um valor fixo de iterações com base na quantidade de conceitos de cada conjunto de Conceitos Afins. Foi utilizado duas vezes esta quantidade.

Assim, este módulo fará a análise dos sub-elementos *relatesSourceEntity* e *relatesTargetEntity* do elemento *ConceptBridge*. As figuras a seguir (de 5.8 a 5.12), apresentam informações sobre os mapeamentos entre os conceitos das diversas ontologias, CPAD, LEGISMAR, SIGDEM e PESSOAL.

```
...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file://D:/MAPS/CPAD.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file://D:/MAPS/LEGISMAR.rdfs"/>
4 </a:Mapping>
5 <a:ConceptBridge rdf:ID="i-1193497595578-809857777" rdfs:label="">
6   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
7   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#DOCUMENTOS"/>
8 </a:ConceptBridge>
9 <a:ConceptBridge rdf:ID="i-1193497715203-592908029" rdfs:label="">
10  <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/CPAD#OM"/>
11  <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
12 </a:ConceptBridge>
...
```

FIG. 5.8 Exemplo de mapeamentos entre conceitos - ponte semântica (*ConceptBridge*)

```
...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file://D:/MAPS/CPAD.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file://D:/MAPS/PESSOAL.rdfs"/>
4 </a:Mapping>
5 <a:ConceptBridge rdf:ID="i-1193501224328-510816776">
6   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
7   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/PESSOAL#DOCUMENTOS"/>
8 </a:ConceptBridge>
...
```

FIG. 5.9 Exemplo de mapeamentos entre conceitos - ponte semântica (*ConceptBridge*)


```

...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file:/D:/MAPS/LEGISMAR.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file:/D:/MAPS/CPAD.rdfs"/>
4 </a:Mapping>
5 <a:ConceptBridge rdf:ID="i-1193416626734-1516085849">
6   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#DOCUMENTOS"/>
7   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
8 </a:ConceptBridge>
9 <a:ConceptBridge rdf:ID="i-1193505539671-650514992">
10  <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
11  <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/CPAD#OM"/>
12 </a:ConceptBridge>
...

```

FIG. 5.10 Exemplo de mapeamentos entre conceitos - ponte semântica (*ConceptBridge*)

```

...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file:/D:/MAPS/PESSOAL.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file:/D:/MAPS/CPAD.rdfs"/>
4 </a:Mapping>
5 <a:ConceptBridge rdf:ID="i-1188061548296-1536185278">
6   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/PESSOAL#DOCUMENTOS"/>
7   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
8 </a:ConceptBridge>
...

```

FIG. 5.11 Exemplo de mapeamentos entre conceitos - ponte semântica (*ConceptBridge*)

```

...
1 <a:Mapping rdf:ID="">|
2   <a:relatesSourceOntology rdf:resource="file:/D:/MAPS/SIGDEM.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file:/D:/MAPS/LEGISMAR.rdfs"/>
4 </a:Mapping>
5 <a:ConceptBridge rdf:ID="i-1193415328046-1301973541">
6   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/SIGDEM#EXPEDIENTE"/>
7   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#DOCUMENTOS"/>
8 </a:ConceptBridge>
9 <a:ConceptBridge rdf:ID="i-1193415328046-1131069873">
10  <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/SIGDEM#ORGAO"/>
11  <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
12 </a:ConceptBridge>
...

```

FIG. 5.12 Exemplo de mapeamentos entre conceitos - ponte semântica (*ConceptBridge*)

Para verificar os conceitos que irão compor a OE, as definições básicas 16, 19 e a heurística 2 do capítulo 4 foram utilizadas.

Inicialmente, será apresentado o passo a passo de como foi realizada a representação dos mapeamentos entre conceitos através de grafos, e sua separação em sub-grafos, em virtude das definições 16 (Conceitos Afins) e 19 (Ranking de um conceito) serem baseadas neste tipo de abordagem.

Novamente, uma matriz de adjacência (simétrica) foi utilizada para a representação do grafo com os diversos mapeamentos entre conceitos. Esta matriz, denominada de Matriz de Mapeamentos de Conceitos, possui igualmente uma coluna com os nomes dos conceitos das ontologias fonte e uma linha com os nomes dos conceitos das ontologias alvo. O valor

verdadeiro (*true*) é atribuído ao par (coluna, linha) caso existam mapeamentos entre estes conceitos, e falso (*false*) em caso contrário.

A construção de um grafo com os mapeamentos dos conceitos, ilustrados nos exemplos das figuras de 5.8 a 5.12 é representado na Matriz de Mapeamentos de Conceitos da tabela 5.2.

Conceitos Abto ↓	Conceitos Fonte →	CPAD #DOCUMENTOS	CPAD #OM	LEGISMAR #DOCUMENTOS	LEGISMAR #OM	PESSOAL #DOCUMENTOS	SIGDEM #EXPEDIENTE	SIGDEM #ORGAO
CPAD#DOCUMENTOS		<i>False</i>	<i>False</i>	<i>True</i>	<i>False</i>	<i>True</i>	<i>False</i>	<i>False</i>
CPAD#OM		<i>False</i>	<i>False</i>	<i>False</i>	<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>
LEGISMAR#DOCUMENTOS		<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>	<i>False</i>	<i>True</i>	<i>False</i>
LEGISMAR#OM		<i>False</i>	<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>	<i>False</i>	<i>True</i>
PESSOAL#DOCUMENTOS		<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>	<i>False</i>	<i>False</i>	<i>False</i>
SIGDEM#EXPEDIENTE		<i>False</i>	<i>False</i>	<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>	<i>False</i>
SIGDEM#ORGAO		<i>False</i>	<i>False</i>	<i>False</i>	<i>True</i>	<i>False</i>	<i>False</i>	<i>False</i>

TAB. 5.2 Matriz de adjacência representativa dos mapeamentos entre conceitos existentes em um ponto

Tendo os mapeamentos entre os conceitos das ontologias sido representados através desta matriz, o sistema irá separar os sub-grafos conexos deste grafo. Estes sub-grafos correspondem aos conjuntos de Conceitos Afins. Para obtenção destes conjuntos, foi aplicada a busca em profundidade na matriz, visando auxiliar a separação de cada parte conexa do grafo (definição básica 16). Identificados os conjuntos de Conceitos Afins, o sistema aplicará a definição básica 19 (*Ranking* de um conceito) e em seguida aplicará a heurística 2 (Eleição de Conceitos para a OE), a fim de selecionar, dentre os Conceitos Afins, o conceito que será representado na OE. A FIG. 5.13 apresenta o grafo formado pelos conceitos e seus mapeamentos representados pela matriz de Mapeamentos de Conceitos. Esta matriz apresenta todos os mapeamentos, não levando em consideração a direção dos mesmos. Por exemplo, o conceito SIGDEM#ORGAO foi mapeado para o conceito LEGISMAR#OM. O mapeamento inverso, isto é, de LEGISMAR#OM para SIGDEM#ORGAO não existe de fato, apesar de estar representado na matriz. Esta matriz foi montada apenas com o objetivo de facilitar a separação das partes conexas. Neste exemplo, a matriz de mapeamentos entre conceitos possui dois sub-grafos conexos, semelhantes aos apresentados na FIG. 5.13a e FIG. 5.13b, com somente uma aresta interligando os vértices.

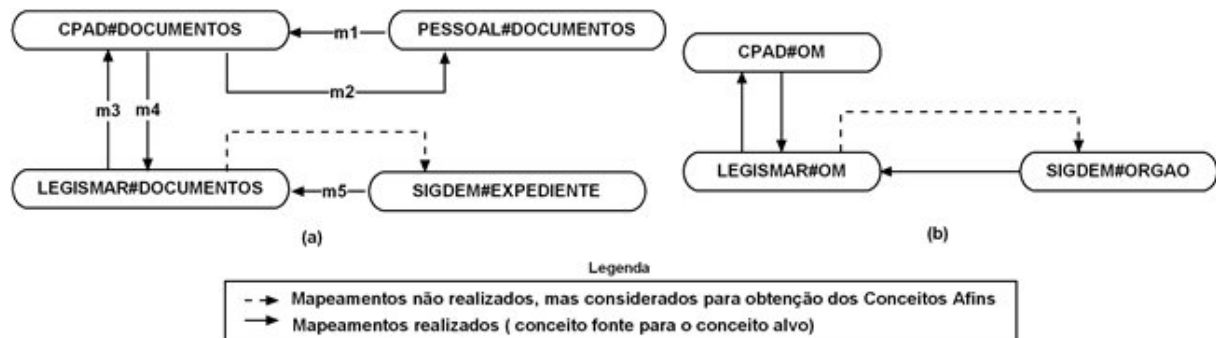


FIG. 5.13 Representação dos mapeamentos entre conceitos através de um grafo

Após identificar os Conceitos Afins, o módulo de Geração da OE prossegue, já podendo calcular o *ranking* dos conceitos para o conjunto de Conceitos (FIG. 5.13a e FIG. 5.13b), utilizando a definição básica 19 (*Ranking* de um conceito). A expressão de cálculo do *ranking* utiliza o grau de saída do vértice e a pontuação dos mapeamentos. Sendo assim, foi montada uma matriz de adjacência, considerando a direção dos mapeamentos dos conceitos. Pelo exemplo anterior, tem-se a matriz que se segue, visualizada na tabela 5.3, os valores da pontuação substituem o valor *true*, que neste exemplo consiramos o valor da pontuação igual a 1, e para os conceitos que não foram mapeados, valor *false*, na tabela anterior, com o valor

da pontuação igual a 0. Note que os mapeamentos inversos, que não existem de fato, não foram representados nesta matriz (tabela 5.3).

Conceitos Abx ↓	Conceitos Fonte →	CPAD #DOCUMENTOS	CPAD #OM	LEGISMAR #DOCUMENTOS	LEGISMAR #OM	PESSOAL #DOCUMENTOS	SIGDEM #EXPEDIENTE	SIGDEM #ORGÃO
CPAD#DOCUMENTOS		0	0	1	0	1	0	0
CPAD#OM		0	0	0	1	0	0	0
LEGISMAR#DOCUMENTOS		1	0	0	0	0	1	0
LEGISMAR#OM		0	1	0	0	0	0	1
PESSOAL#DOCUMENTOS		1	0	0	0	0	0	0
SIGDEM#EXPEDIENTE		0	0	0	0	0	0	0
SIGDEM#ORGÃO		0	0	0	0	0	0	0

TAB 5.3 Matriz de adjacência representativa dos mapeamentos entre conceitos existentes em um ponto, considerando a direção dos mapeamentos

Em resumo, foram criadas duas matrizes de mapeamentos de conceitos, descritas nas tabelas 5.2 e 5.3, sendo que uma delas considera a direção do mapeamento (tabela 5.3).

A matriz representada pela tabela 5.2 é utilizada apenas para a separação dos conjuntos de Conceitos Afins, enquanto que a matriz representada na tabela 5.3 é usada para calcular o o *ranking*, definição básica 19, usando as informações do grau de saída de cada vértice e a pontuação do mapeamento.

Depois de calculados os *rankings* de todos os conceitos (definição básica 19), o sistema aplicará a heurística 2, que visa selecionar o conceito de maior *ranking* entre os Conceitos Afins.

No exemplo da FIG. 5.13a e FIG. 5.13b, os conceitos CPAD#DOCUMENTOS e LEGISMAR#OM, foram os conceitos de maior ranking e assim identificados pela heurística 2, como Conceitos Eleitos (CE). Estes conceitos passam então a compor a OE. A seguir, o sistema passará à seleção das propriedades que deverão compor a OE e interligar seus conceitos.

Para seleção das propriedades o sistema não consultará os arquivos de mapeamentos. Para cada conceito do conjunto CE (definição básica 17), serão aplicadas as heurísticas 3 e 4. Para isso, o sistema verificará, no arquivo onde é descrita a ontologia (RDF) da qual faz parte cada conceito, identificando quais as propriedades literais (heurística 3) e quais os relacionamentos (heurística 4) que partem deste conceito.

No exemplo utilizado têm-se $CE = \{CPAD\#DOCUMENTOS, LEGISMAR\#OM\}$, isto é, os dois conceitos eleitos para serem representados na OE.

O conceito DOCUMENTOS da ontologia CPAD (FIG. 5.14) contém a propriedade ASSUNTO e os relacionamentos DESTINO e ORIGEM, que possuem como *range* o conceito CPAD#OM. No entanto, o conceito CPAD#OM não pertence a CE, isto é, não é um conceito eleito, porém foi um conceito candidato (definição básica 15), pois teve afinidade com o conceito LEGISMAR#OM, como mostra na FIG. 5.13b. Então o range do relacionamento será modificado para LEGISMAR#OM pelo fato deste último pertencer a CE.

Como o conceito OM pertence à ontologia LEGISMAR (FIG. 5.15), então sua propriedade NOME também aparecerá na OE.

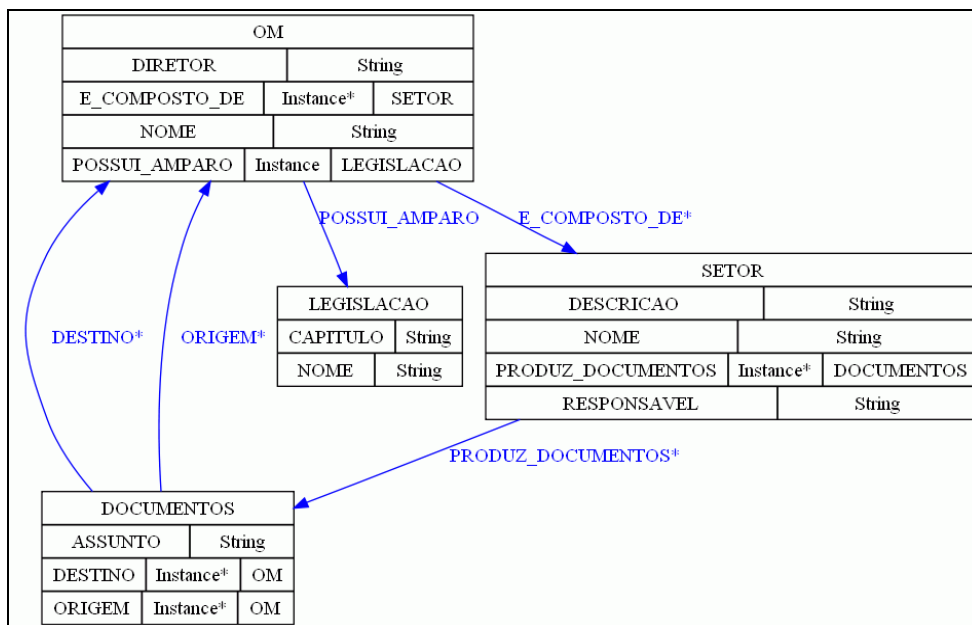


FIG. 5.14 Extrato da ontologia CPAD

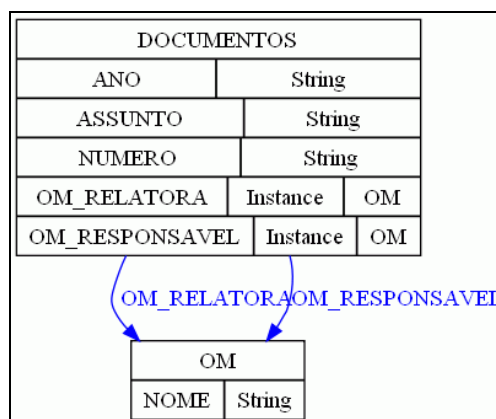


FIG. 5.15 Extrato da ontologia LEGISMAR

Por fim, a representação final da OE é apresentada na FIG. 5.16 graficamente e na FIG. 5.17 em RDF.

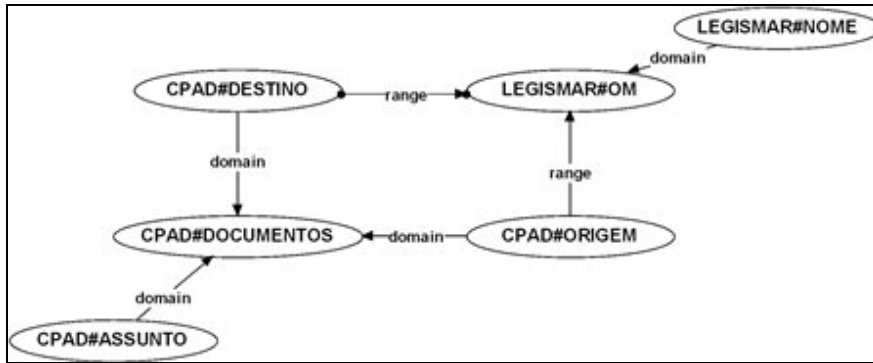


FIG. 5.16 Representação final da OE

```

<rdf:RDF
  xmlns="http://www.owl-ontologies.com/EMERGE_ONTO.owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:p1="http://protege.stanford.edu/CPAD#"
  xmlns:p2="http://protege.stanford.edu/LEGISMAR#"
  xmlns:daml="http://www.daml.org/2001/03/daml+oil#">

  <rdf:Description rdf:about="http://www.owl-ontologies.com/EMERGE_ONTO.owl#"
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Ontology"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/CPAD#DOCUMENTOS">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/CPAD#ASSUNTO">
    <rdfs:domain rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/CPAD#DESTINO">
    <rdfs:range rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
    <rdfs:domain rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/CPAD#ORIGEM">
    <rdfs:range rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
    <rdfs:domain rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/LEGISMAR#OM">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/LEGISMAR#NOME">
    <rdfs:domain rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
  </rdf:Description>
</rdf:RDF>

```

FIG. 5.17 Representação final da OE em RDF

5.2.4 MÓDULO DE COMUNICAÇÃO E GERÊNCIA PONTO-A-PONTO

Como foi dito anteriormente, o módulo de Gerência avalia se existem mapeamentos entre ontologias em um determinado ponto. Os mapeamentos devem englobar todos os pontos que possuem ontologias e que utilizam o sistema *OntoEmerge*.

Para que os mapeamentos realizados por esses pontos se tornem disponíveis para avaliação do módulo de Gerência, transferências dos arquivos de mapeamentos são realizadas entre os diversos pontos. Por exemplo, na FIG. 4.6, existe um mapeamento realizado entre a ontologia do ponto X com a ontologia do ponto Y. Após o mapeamento, o arquivo (SBO/RDF) resultante será transferido para o ponto Y e, caso haja outros mapeamentos, por exemplo, do ponto Y com o ponto J, este processo se repete. No momento em que o módulo de Gerência avalia que em determinado ponto existem mapeamentos envolvendo todos os pontos, ele invoca o módulo de Geração da OE.

O módulo também disponibiliza, para todos os pontos, cópia das ontologias existentes dos diversos pontos, realizando transferências destes arquivos, e ainda os arquivos de suas instâncias (RDF).

Os arquivos de mapeamentos, ontologias e instâncias são armazenados no BD MAP, BD ONTOLOGIAS e BD INSTANCIAS, respectivamente. Estas bases de dados são diretórios do sistema operacional. Assim, através da transferência destes arquivos, estes passam a ser compartilhados pelos diversos pontos, usuários do *OntoEmerge*.

A transferência dos arquivos supracitados é realizada de forma transparente para o usuário, através do módulo de Comunicação e Gerência P2P. A funcionalidade para a transferência de arquivos foi desenvolvida utilizando-se a plataforma COPPEER (MIRANDA, 2006), que provê outras funcionalidades, tais como: gerência e comunicação entre os pontos e outras pertinentes à camada de rede.

5.2.4.1 COPPEER - INFRA-ESTRUTURA PARA GERÊNCIA P2P (MIRANDA, 2006)

O COPPEER é uma plataforma que se destina à implementação de aplicações colaborativas em ambientes P2P, em desenvolvimento pela COPPE Sistemas da UFRJ. Possui uma arquitetura lógica de vários níveis, destinada a facilitar o tratamento da complexidade combinada de sistemas P2P e de colaboração. A FIG. 5.18 apresenta a arquitetura lógica da COPPEER. Seus módulos são descritos a seguir:



FIG. 5.18 Arquitetura lógica da plataforma COPPEER (MIRANDA, 2006)

- **Camada de agência:** executa computações P2P genéricas para atender as necessidades das outras camadas. Dentro dessa camada, o sistema é visto como um conjunto de agências colaborativas. Uma agência é a entidade interna de um nó da rede onde os agentes do Coppeer são executados. Um agente Coppeer é um artefato de *software* que troca informações com outros agentes, movendo-se de uma agência para outra para executar computações distribuídas. Esta camada foi objeto de pesquisa que resultou na implementação de um ambiente em tempo de execução para aplicações P2P denominado CoppeerCAS, que oferece suporte para a aplicação de princípios de Sistemas Complexos Adaptativos (*Complex Adaptive Systems*);
- **Camada de integração:** tem como objetivo resolver problemas de locação de aplicações, gerenciamento de sessões e integração remota de sistemas P2P. É também responsável por fornecer locações virtuais de alta disponibilidade para aplicações colaborativas. Como as agências são dependentes dos nós da rede, que são a princípio transientes, entidades virtuais de alta disponibilidade devem ser

mantidas coletivamente por conjuntos de agências, de forma a assegurar o suporte das informações das sessões de trabalho colaborativo. Simplifica o desenvolvimento de aplicações, provendo abstrações de interação entre entidades localizadas em nós diferentes do sistema, não exigindo do usuário, o conhecimento sobre a infra-estrutura de rede utilizada; e

- **Camada de colaboração:** apóia-se na camada de integração para implementar mecanismos de alto nível para cooperação, coordenação e comunicação entre usuários.

A escolha desta plataforma para a gerência do ambiente P2P se atribui à facilidade de acesso ao ambiente e acesso a documentação, em virtude da disponibilidade e proximidade de contato com seus desenvolvedores. Outras abordagens poderiam ser utilizadas para implementação deste módulo, conforme apresentadas no capítulo 3.

5.2.4.2 IMPLEMENTAÇÃO DO MÓDULO DE COMUNICAÇÃO E GERÊNCIA P2P

A aplicação criada é registrada em um arquivo de configuração, próprio do COPPEER, a fim de que seja provida instância e um item de menu correspondente em sua janela principal para que os usuários possam invocá-la.

Conforme mostrado na FIG. 5.19, foi criada uma aplicação chamada *TransferFiles*, que realizará a transferência de arquivos entre os diversos pontos.

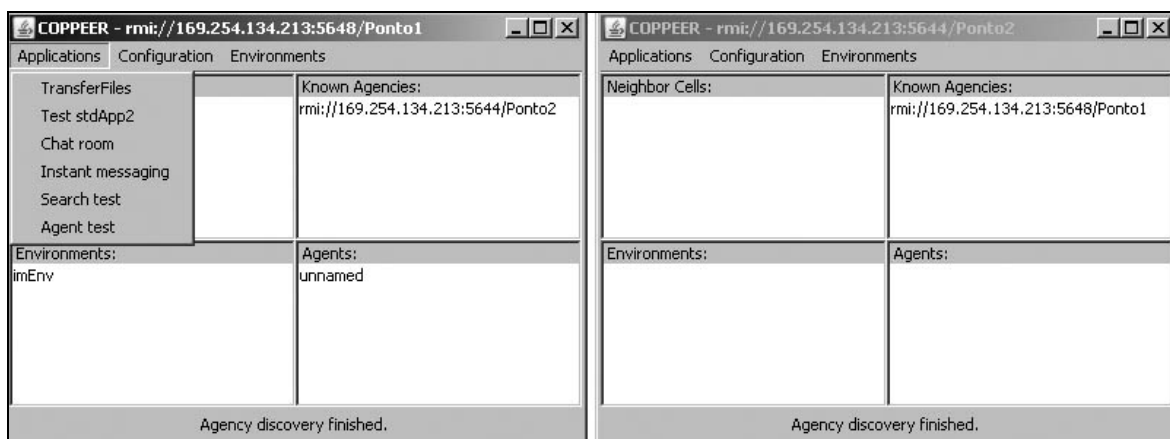


FIG. 5.19 Janela principal do COPPEER

As funcionalidades adicionais para este módulo tiveram como base o exemplo existente no COPPEER, e são descritas a seguir:

- *Enviar Arquivos* - após salvar os arquivos da ontologia do ponto, suas instâncias e/ou os mapeamentos realizados, o módulo automaticamente envia aos demais pontos estas informações e aquelas, que por ventura, já estejam nas bases de dados (BD MAP, BD INSTANCIAS e BD ONTOLOGIAS); e
- *Armazenar Arquivos* - os vários arquivos, encaminhados por outros pontos são recebidos e armazenados nas bases de dados. Para que o sistema armazene corretamente estes arquivos, uma seleção dos arquivos é realizada, previamente, de acordo com sua extensão ou o próprio nome dos arquivos que pode obedecer algum padrão. Por exemplo, no sistema proposto, os mapeamentos possuem extensão ".rdf", sendo assim, serão armazenados no BD MAP.

Alguns testes de transferência de arquivos, de tamanhos entre 1Kb e 64Mb, foram realizados, localmente, e não apresentaram problemas.

5.3 CONSIDERAÇÕES FINAIS

Este capítulo tratou do desenvolvimento do protótipo *OntoEmerge*, sendo que suas funcionalidades serão melhor elucidadas com a utilização de exemplos. As ontologias usadas como exemplo serão descritas em mais detalhes no próximo capítulo. Da arquitetura proposta, nem todas as funcionalidades do protótipo foram desenvolvidas, porém o objetivo de gerar uma visão global, aqui denominada OE, utilizando os mapeamentos disponíveis nos diversos pontos foi alcançado.

O módulo de Geração da OE foi desenvolvido, enquanto que o módulo Gerência foi parcialmente implementado. A ferramenta MAFRA foi utilizada com o suporte para resolver as funcionalidades do módulo de Mapeamento P2P. O módulo de Comunicação e Gerência P2P foi implementado tendo como suporte a ferramenta COPPEER (MIRANDA, 2006), que proporciona toda infra-estrutura para rede P2P, como por exemplo: gerência e comunicação entre os pontos. A funcionalidade básica desenvolvida com os recursos do COPPEER foi a de transferência de arquivos entre pontos.

Foram utilizados grafos para representação dos mapeamentos entre ontologias e entre conceitos, sendo que a estrutura de dados usada foi a matriz de adjacência. A busca em profundidade na matriz proporcionou a verificação da conexidade do grafo, implicando diretamente na definição do momento correto para geração da OE e na separação dos Conceitos Afins.

No próximo capítulo será apresentado um exemplo de uso que visa elucidar com maior clareza o passo-a-passo de utilização do protótipo, bem como avaliar o sistema através de testes exaustivos, demonstrando a representatividade da OE diante dos mapeamentos e ontologias utilizadas.

6 EXEMPLO DE USO E AVALIAÇÃO DO PROTÓTIPO

Este capítulo apresenta duas iniciativas deste trabalho no sentido de avaliar a proposta apresentada no capítulo 4. A primeira delas é um exemplo de uso, onde serão utilizadas quatro ontologias, com o objetivo de mostrar o funcionamento do sistema. A segunda será a realização de testes que farão uso de trinta e uma ontologias propostas pela *Ontology Alignment Evaluation Initiative - OAEI* (OAEI), visando obter informações quantitativas das OEs geradas e dos mapeamentos utilizados.

O exemplo de uso inicia-se na seção 6.1 e nos itens seguintes são apresentadas a construção das ontologias utilizadas, os mapeamentos realizados, a análise dos mapeamentos e a geração da OE. A seção 6.2 apresenta o detalhamento da bancada de testes desenvolvida para avaliação da ferramenta com realização de alguns experimentos. Por fim, na seção 6.3 as considerações finais do capítulo.

6.1 EXEMPLO DE USO DO PROTÓTIPO ONTOEMERGE

O Exemplo de uso escolhido teve como contexto o cenário de sistemas de informação em uso atualmente pela Marinha, que tratam de *workflow*, arquivamento e legislação de documentos.

Em organizações de médio e grande porte, como a Marinha, vários sistemas de informação são desenvolvidos em paralelo, e sem necessariamente seguir uma terminologia ou conceituação comum. Na maioria das vezes isso se deve ao fato da organização não disponibilizar uma ontologia organizacional representativa. Assim, além da heterogeneidade encontrada com a plataforma operacional, linguagem de implementação, sistema de gerenciamento de banco de dados, e ainda esquema de dados, a interoperabilidade entre estes sistemas é prejudicada também pela heterogeneidade semântica.

Os diversos sistemas existentes, em sua maioria, possuem modelos que representam dados e informações de processos que foram utilizados no seu desenvolvimento e que podem ainda ser utilizados por um modelo organizacional. O principal objetivo de um modelo

organizacional é obter uma visão da organização, que pode ser utilizada para diversos fins. Tem-se defendido na literatura (ALMEIDA, 2006) que ontologias são estruturas adequadas para a criação de modelos organizacionais, pois possuem conceitos, relações e atributos semanticamente bem definidos, sendo que a linguagem utilizada para sua construção pode variar em grau de formalidade, conforme a necessidade. Além disso, a linguagem formal das ontologias, isenta de ambigüidades, permite que estas sejam processáveis por máquina.

Dessa forma, o exemplo de uso apresentado nesta seção tem como objetivo ilustrar a proposta deste trabalho no contexto de alguns sistemas usados na Marinha, inserida no domínio de documentos e na tarefa de controle de documentos, pressupondo o uso de quatro ontologias, e gerando uma visão global inicial. A construção de cada ontologia é discutida na subseção 6.1.1 e os mapeamentos realizados entre as mesmas com a finalidade de interoperar são apresentados na subseção 6.1.2.

Dentre as ontologias construídas, três representam sistemas de informação existentes na Marinha e uma quarta ontologia foi construída para representar a visão e o conhecimento de alguns usuários da organização. A metodologia de construção das ontologias foi baseada em algumas abordagens (NOY, et al., 2001), (USCHOLD, et al., 1995), (GUARINO, 1997), (FERNANDEZ, et al., 1997), (GRUBER, 1995) e (GRÜNINGER & FOX, 1995) encontradas na literatura e em uma visão subjetiva do domínio, com a finalidade de, somente, ilustrar, num contexto real, a aplicabilidade da proposta.

A linguagem utilizada para expressar as ontologias foi a RDFS (RDFS), mas para facilitar a compreensão do leitor, as ontologias são apresentadas graficamente.

Depois de construídas as ontologias, diversos mapeamentos foram realizados entre elas, fazendo uso da ferramenta MAFRA. Estes mapeamentos foram então submetidos ao sistema *OntoEmerge*, gerando uma ontologia final, que seria a representação global das quatro ontologias utilizadas, ou seja, a visão global inicial.

6.1.1 CONSTRUÇÃO DAS ONTOLOGIAS

Como dito anteriormente, a construção das ontologias utilizadas neste exemplo de uso, não seguiu metodologias específicas como as propostas na literatura, mas algumas de suas tarefas e recomendações relevantes foram usadas. Assim, foram realizadas visitas a Organizações

Militares (OMs) da Marinha com atividades fim diferentes, onde foram desenvolvidos os sistemas em uso, selecionados para este estudo. No processo de construção levou-se em consideração quatro tarefas, descritas a seguir:

- *Entrevistas com especialistas do domínio:* visam obter conhecimento sobre as características do domínio, de forma a limitar o seu escopo;
- *Levantamentos de fontes:* teve com o objetivo analisar relatórios e sistemas utilizados em cada OM;
- *Levantamento dos termos representativos dos conceitos relevantes:* visou gerar um mapa conceitual, organizando os termos encontrados. Para isto, foi utilizada a ferramenta JUDE (JUDE) para edição do mapa conceitual; e
- *Formalização da ontologia:* foi realizada através da linguagem de ontologias RDFS (RDFS), com a utilização da ferramenta de edição de ontologias Protégé (PROTÉGÉ). Porém, para facilitar a compreensão serão apresentadas graficamente.

6.1.1.1 ONTOLOGIA CPAD

Esta ontologia foi construída com base no sistema *Comissão Permanente de Avaliação de Documentos (CPAD)*. Este sistema tem por finalidade, auxiliar na definição da temporalidade de arquivamento de documentos da Marinha. Foram extraídos os conceitos principais, conforme a FIG. 6.1 e destes foram escolhidos alguns para compor a ontologia, representada graficamente em RDF, como mostra a FIG. 6.2.

Toda Organização Militar (OM) possui setores, os quais produzem documentos e são enviados ou recebidos para/de outras OMs. Devido à necessidade de definir o tempo que um documento deva permanecer arquivado e/ou guardado, foi desenvolvido um sistema para sua classificação. A definição do prazo para arquivamento vai depender de sua espécie, que é caracterizada, principalmente por:

- Arquivo corrente: prazo do arquivo de primeira idade e que tem sob sua guarda documentos freqüentemente consultados e que cumprem ainda as finalidades que motivaram a sua produção;
- Arquivo Intermediário: prazo do arquivo de segunda idade e que tem sob sua guarda documentos de uso eventual pelo setor que os produziu, devendo ser conservados em

depósitos de armazenagem temporária, aguardando sua eliminação ou recolhimento para guarda permanente; e

- Destino final: após o cumprimento do prazo de guarda no arquivo corrente, o documento pode ser transferido para o arquivo intermediário ou ser eliminado. Após o cumprimento do prazo de guarda no arquivo intermediário, o documento pode ser recolhido ao arquivo permanente (preservar), microfilmado ou eliminado.

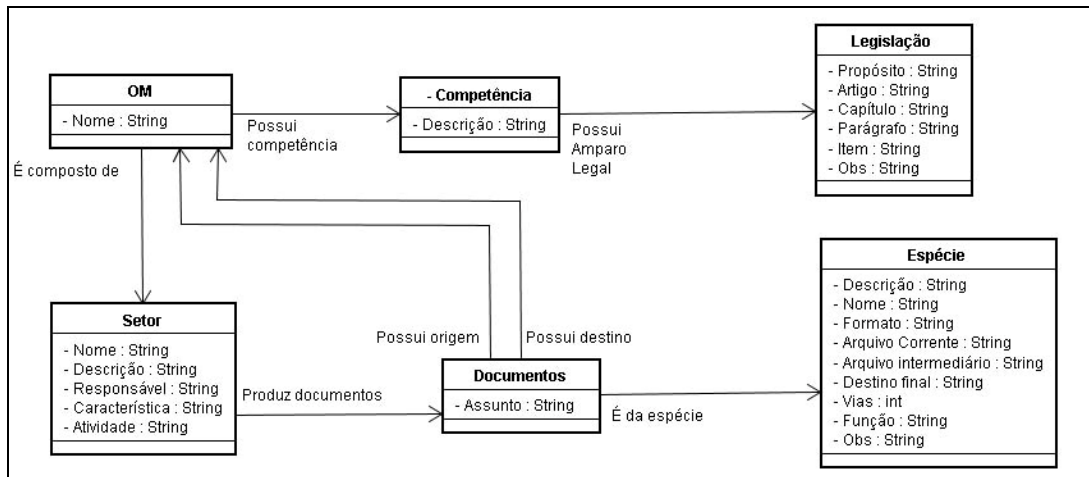


FIG. 6.1 Mapa conceitual do sistema CPAD em UML

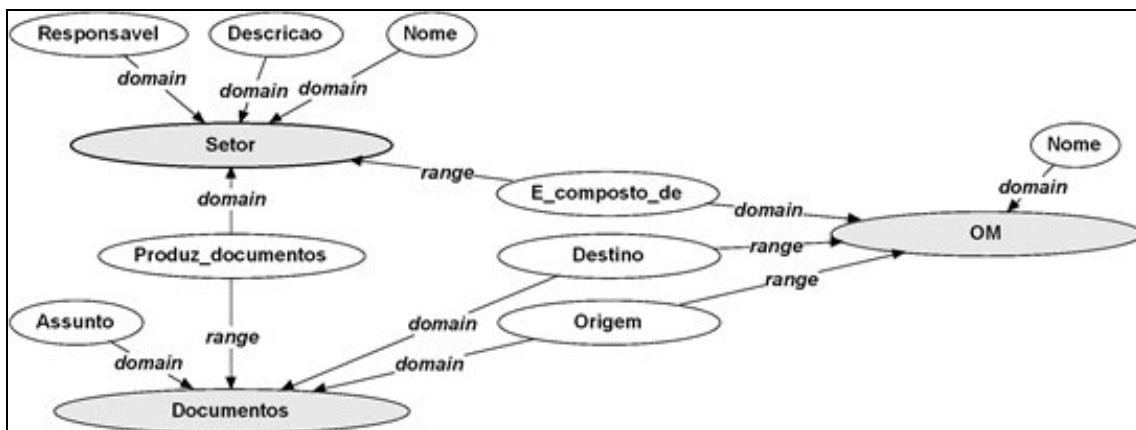


FIG. 6.2 Representação gráfica em RDF da Ontologia CPD

6.1.1.2 ONTOLOGIA LEGISMAR

A ontologia LEGISMAR foi fundamentada no levantamento de informações referentes ao *Sistema Controle de Legislação da Marinha (LEGISMAR)*. Este sistema tem como objetivo permitir a busca e recuperação das legislações de interesse da Marinha. Permite consultas a documentos de forma genérica, boletins, documentos gerados fora do âmbito da Marinha e documentos mais recentes. A FIG. 6.3 apresenta o mapa conceitual do sistema e a FIG. 6.4 a ontologia representada graficamente em RDF.

Os documentos tratados neste sistema são produzidos por uma determinada OM, na qual pode ser responsável e/ou relatora do mesmo. O sistema controla também o armazenamento de seu arquivo, registrando seu nome, nome do pessoa que incluiu este arquivo (Inclusão) e a data que foi incluído.

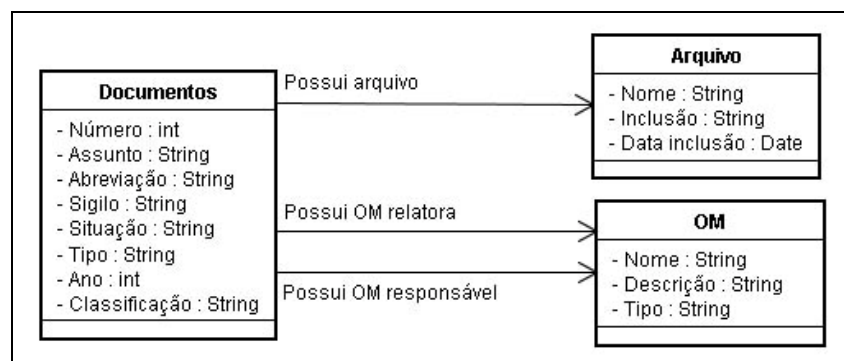


FIG. 6.3 Mapa conceitual do sistema LEGISMAR em UML

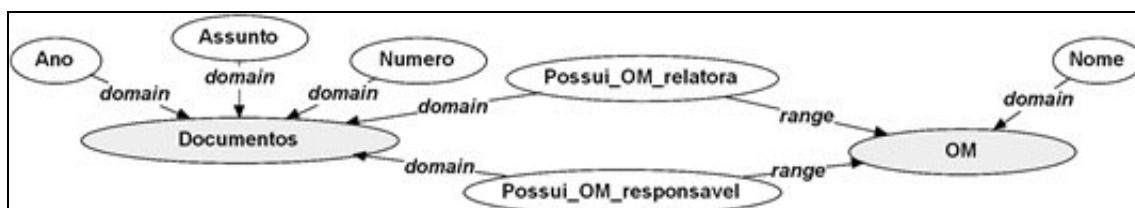


FIG. 6.4 Representação em RDF da Ontologia LEGISMAR

6.1.1.3 ONTOLOGIA SIGDEM

A ontologia SIGDEM teve como base o *Sistema de Gerência de Documentos Eletrônicos da Marinha (SIGDEM)*. Este sistema foi desenvolvido para realizar comunicações entre organizações ou departamentos, com segurança, padronização e rapidez. Suas principais funções são: controle de acesso a documentos, controle do fluxo dos documentos com acompanhamento automático das ações a serem executadas em cada documento, e o acompanhamento do histórico do documento, desde a sua criação ou recebimento. A FIG. 6.5 apresenta o mapa conceitual do sistema e a FIG. 6.6 a ontologia representada graficamente em RDF.

Todo órgão da Marinha é composto por setores, sendo que cada setor possui usuários que acessam diversos expedientes, os quais podem ou não ser enviados para outros setores de diversos órgãos.

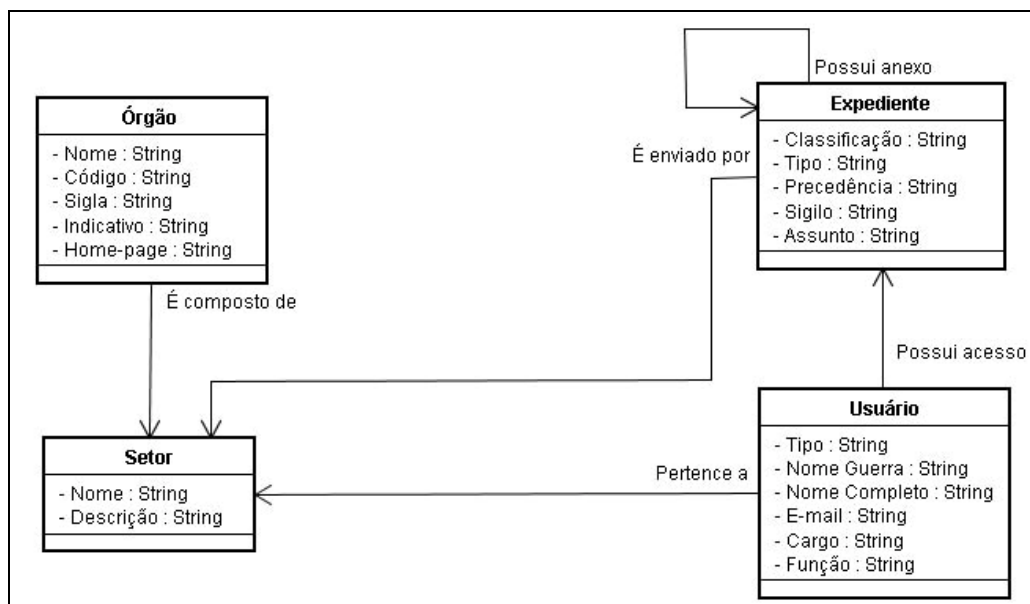


FIG. 6.5 Mapa conceitual do sistema SIGDEM em UML

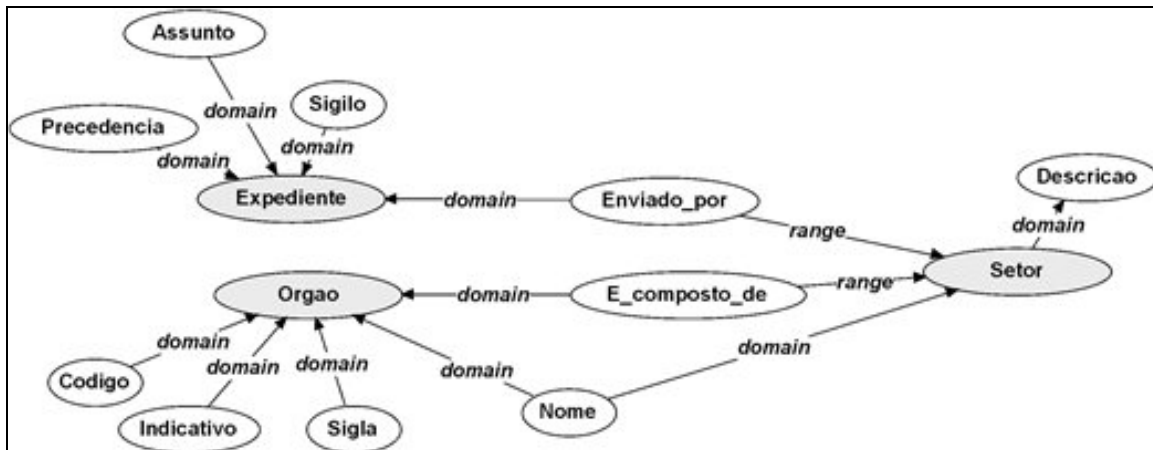


FIG. 6.6 Representação em RDF da Ontologia SIGDEM

6.1.1.4 ONTOLOGIA PESSOAL

Esta ontologia foi construída com base na organização de documentos de três usuários. Tendo em vista que usuários, geralmente, criam diretórios para controle de seus documentos na forma de uma hierarquia de tipos informais (LASSILA & MCGUINNESS, 2001), procurou-se analisar algumas representações e criar uma ontologia. A FIG. 6.7 apresenta o mapa conceitual desta análise e a FIG. 6.8 a ontologia representada graficamente em RDF.

Os documentos são assinados por um determinado servidor publico, sendo que podem fazer referência a outros documentos.

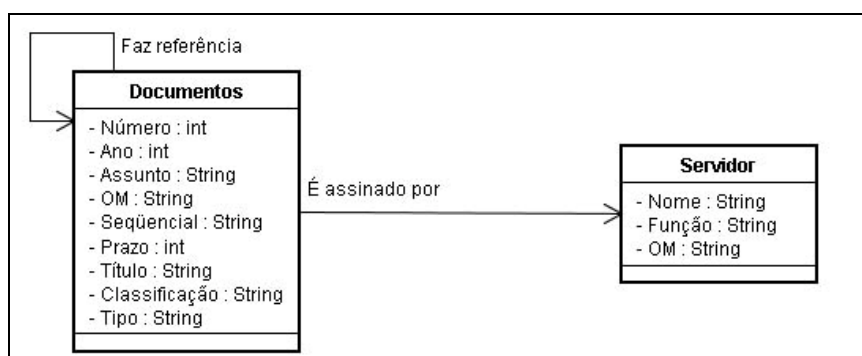


FIG. 6.7 Mapa conceitual do sistema PESSOAL em UML

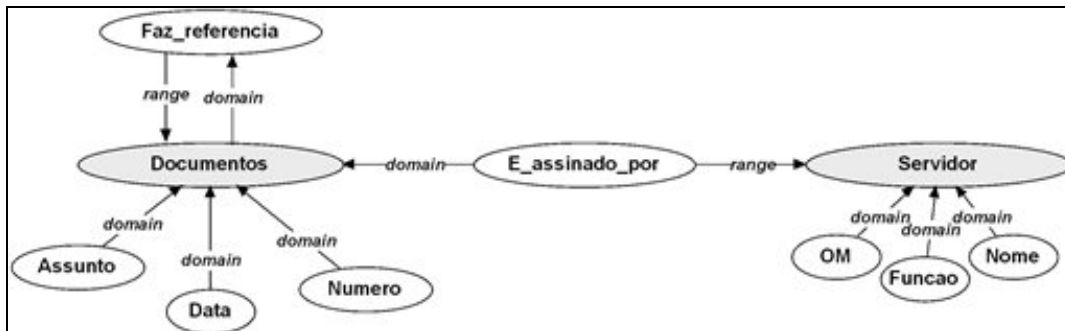


FIG. 6.8 Representação em RDF da Ontologia PESSOAL

6.1.2 MAPEAMENTOS REALIZADOS

Os mapeamentos entre as ontologias foram realizados utilizando-se a ferramenta MAFRA, e representados graficamente em RDF.

Foram realizados 12 mapeamentos com as quatro ontologias construídas sendo que somente dois deles são apresentados graficamente na FIG. 6.12 e FIG. 6.13.

Para a realização de mapeamentos, a ferramenta disponibiliza uma interface gráfica, conforme mostra a FIG. 6.9. Primeiramente cria-se um projeto definindo o seu nome, bem como o diretório onde ele será armazenado. No caso do *OntoEmerge*, deverá ser o diretório que consta no parâmetro de configuração BD-MAP. O usuário deve indicar as ontologias que serão mapeadas, estando estas armazenadas no diretório que consta no parâmetro de configuração BD-ONTOLOGIAS.

A FIG. 6.9 apresenta a tela de configuração do nome e diretório onde será criado o arquivo com os mapeamentos entre as ontologias fonte e alvo.

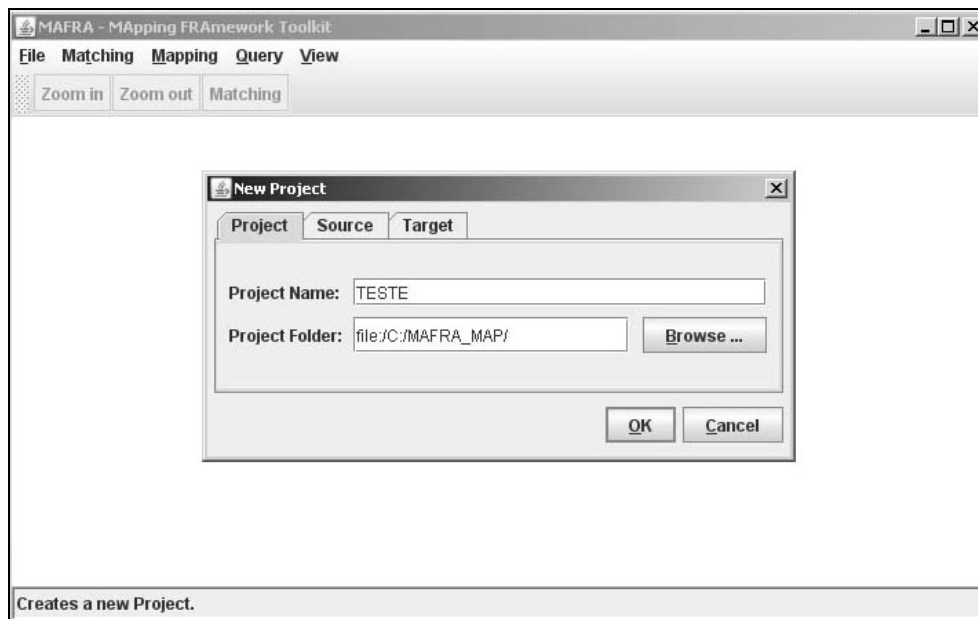


FIG. 6.9 Interface da ferramenta MAFRA para definição do nome do arquivo de mapeamento

Uma vez criado o arquivo de mapeamento, definido pelo nome do projeto, o usuário irá identificar os conceitos, propriedades (relacionamentos) e propriedades (atributos) que poderão ser mapeados, criando as associações (*matches*) correspondentes. Após este passo, a ferramenta MAFRA cria as pontes semânticas necessárias sugerindo alguns serviços de transformações de instâncias.

O MAFRA depende inicialmente e fortemente da criação manual de associações feitas usuário, para estabelecer mapeamentos. Quando estas associações não são definidas, o MAFRA não permite gerar os mapeamentos.

Por exemplo, na FIG. 6.10, tem-se duas ontologias A e B apresentadas pela interface de mapeamento do MAFRA, onde foram associados os conceitos documentos e OM das ontologias e posteriormente criadas as pontes semânticas (FIG. 6.11) representando os conceitos mapeados. Nota-se que a visualização gráfica das ontologias facilita o mapeamento manual.

Todo ponto participante da rede pode realizar mapeamentos de sua ontologia com as ontologias de outros pontos através desta ferramenta.

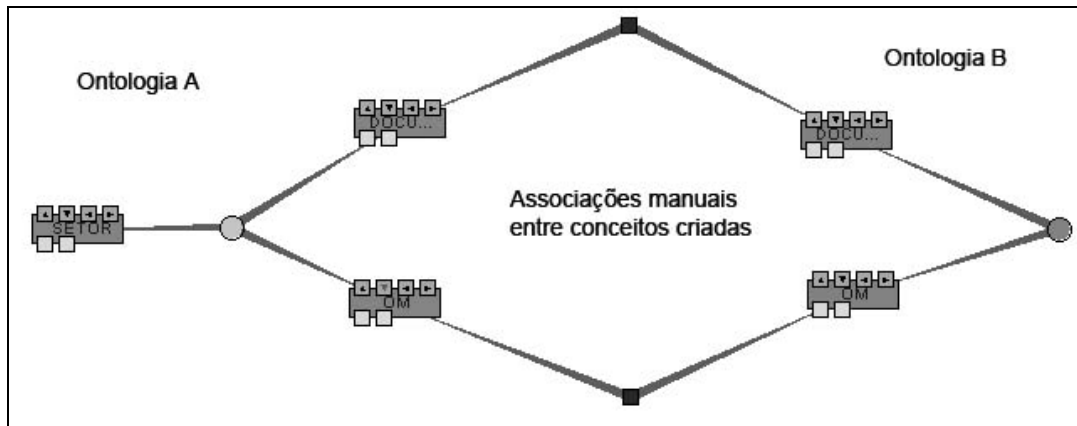


FIG. 6.10 Representação das associações criadas pela interface do MAFRA

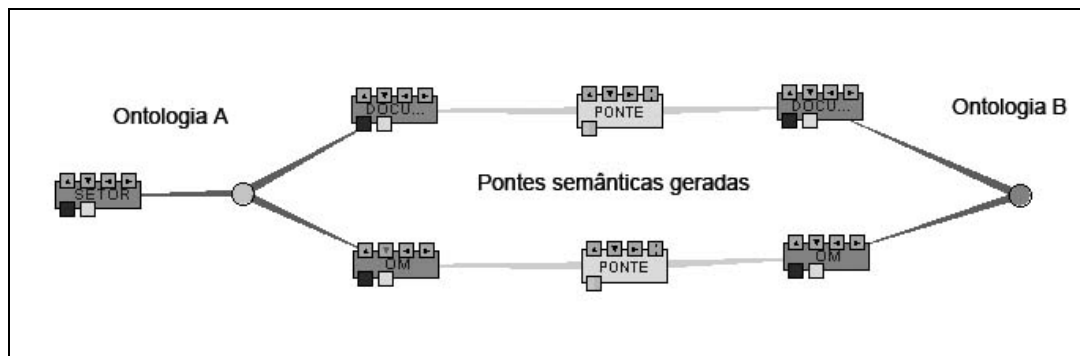


FIG. 6.11 Representação das pontes geradas pela interface do MAFRA

No exemplo da FIG. 6.12 foi realizado o mapeamento entre as ontologias CPAD e LEGISMAR, sendo que os mapeamentos entre conceitos foram:

- a) classe Documentos e seu atributo Assunto da ontologia CPAD com a classe Documento e seu atributo Assunto da ontologia LEGISMAR;
- b) classe Documentos e sua relação Origem da ontologia CPAD com a classe Documentos e sua relação Possui_OM_responsável da ontologia LEGISMAR;
- c) classe OM e seu atributo Nome da ontologia CPAD com a classe OM e seu atributo Nome da ontologia LEGISMAR;

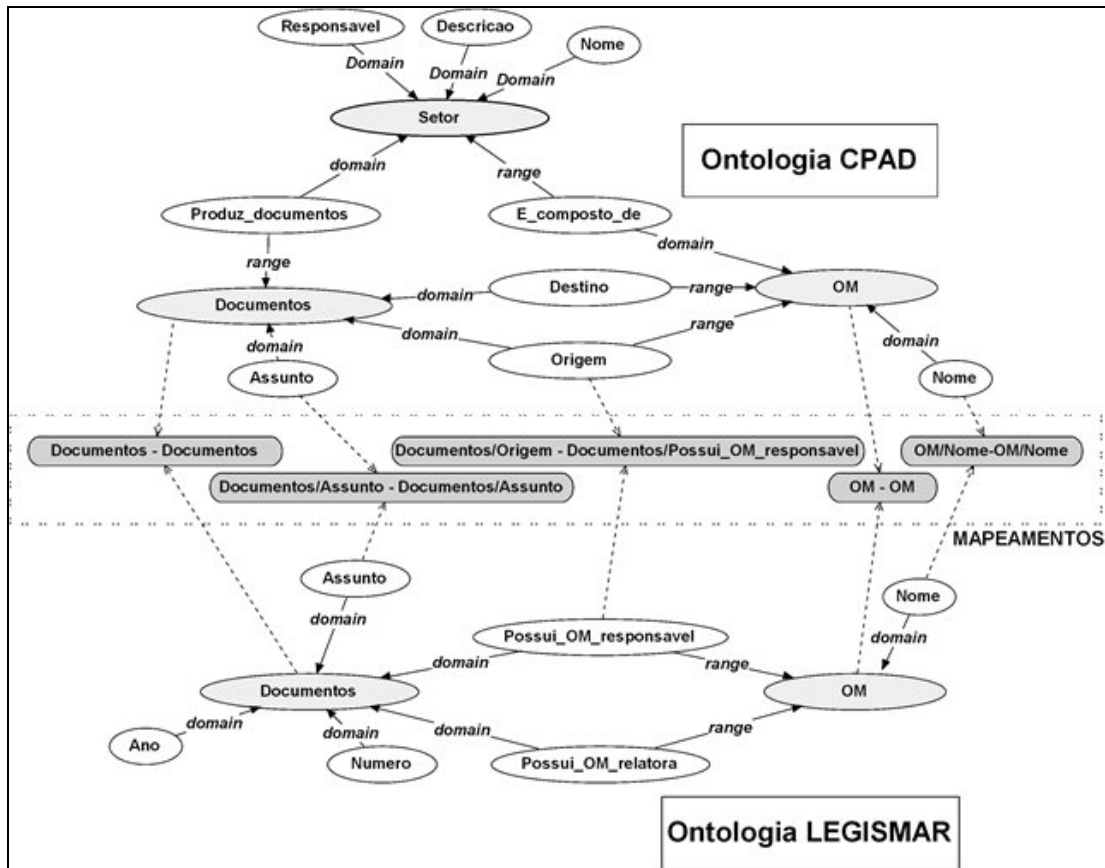


FIG. 6.12 Mapeamento entre a ontologia CPAD e a ontologia LEGISMAR

No exemplo da FIG. 6.13 foi realizado o mapeamento entre as ontologias CPAD e PESSOAL, sendo que os mapeamentos entre conceitos foram:

- classe Documentos e seu atributo Assunto da ontologia CPAD com a classe Documento e seu atributo Assunto da ontologia PESSOAL;
- classe Documentos e sua relação Origem da ontologia CPAD com a classe Documentos e seu atributo OM da ontologia PESSOAL;

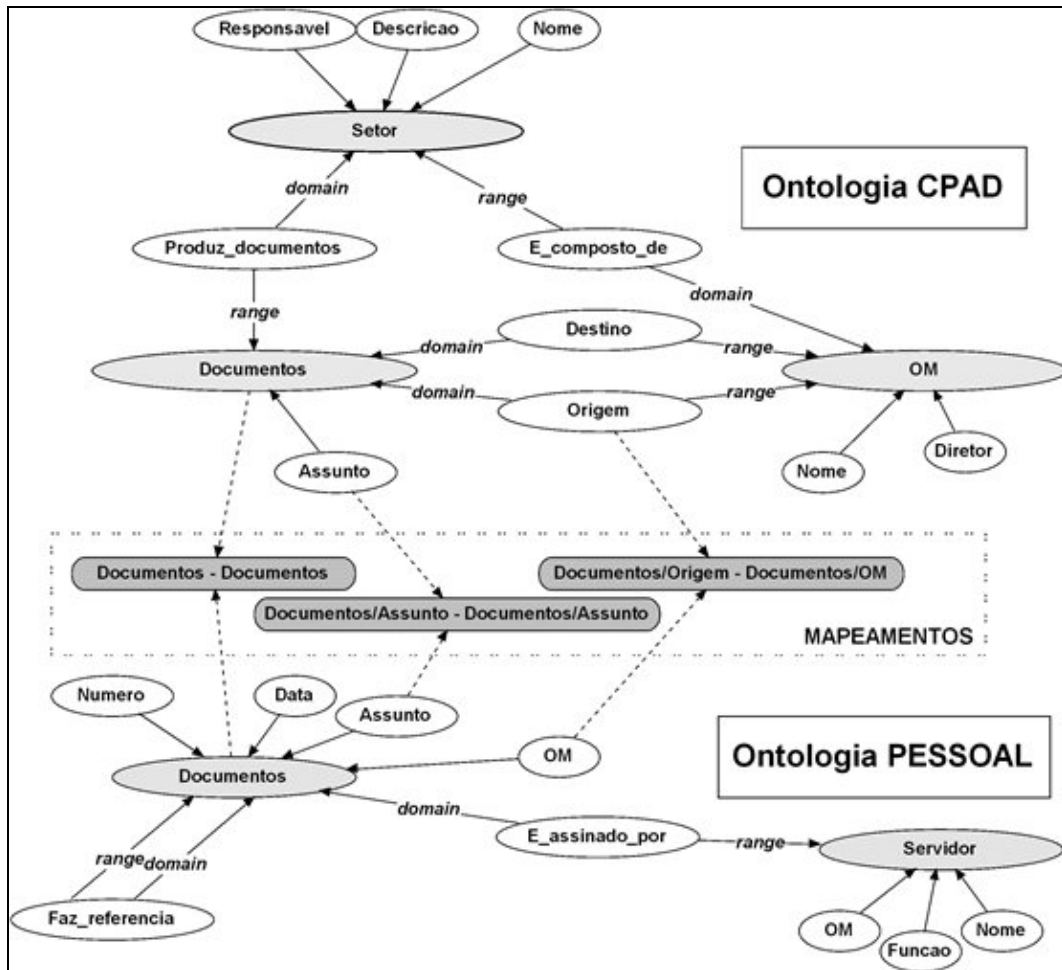


FIG. 6.13 Mapeamento entre a ontologia CPAD e a ontologia PESSOAL

A FIG. 6.14 e FIG. 6.15 apresentam os mapeamentos representados em SBO/RDF (MAEDCHE et al., 2002) onde são destacadas as informações que serão utilizadas pelo *OntoEmerge*. As linhas 2 e 3 da FIG. 6.14 identificam as ontologias que foram mapeadas e as linhas 7, 8, 12 e 13 identificam os conceitos mapeados. Igualmente, na FIG. 6.15, as linhas 2 e 3 identificam as ontologias mapeadas e as linhas 7 e 8 os conceitos mapeados.

```

1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file://D:/MAPS/CPAD.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file://D:/MAPS/LEGISMAR.rdfs"/>
4 </a:Mapping>
5 <a:ConceptBridge rdf:ID="i-1193497595578-809857777"
6   rdfs:label="DOC_CPAD-DOC_LEG">
7   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
8   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#DOCUMENTOS"/>
9 </a:ConceptBridge>
10 <a:ConceptBridge rdf:ID="i-1193497715203-592908029"
11   rdfs:label="CPAD_OM-LEG_OM">
12   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/CPAD#OM"/>
13   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
14 </a:ConceptBridge>
15 </rdf:RDF>

```

FIG. 6.14 Mapeamentos representados em SBO/RDF

```

1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file://D:/MAPS/CPAD.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file://D:/MAPS/PESSOAL.rdfs"/>
4 </a:Mapping>
5 <a:ConceptBridge rdf:ID="i-1193497595578-809857777"
6   rdfs:label="DOC_CPAD-DOC_LEG">
7   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
8   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/PESSOAL#DOCUMENTOS"/>
9 </a:ConceptBridge>

```

FIG. 6.15 Mapeamentos representados em SBO/RDF

6.1.3 ANÁLISE DOS MAPEAMENTOS

Os mapeamentos entre as ontologias CPAD, SIGDEM, PESSOAL e LEGISMAR foram realizados sem dificuldades, em virtude dos mapeamentos não necessitarem de transformações mais complexas.

6.1.4 GERAÇÃO DA ONTOLOGIA EMERGENTE (OE)

Para geração da OE, os dados de entrada compostos pelos arquivos de mapeamentos (Ex. FIG. 6.14 e FIG. 6.15) entre ontologias são monitorados. Este monitoramento ocorre através da verificação da conexidade de um grafo que é construído com as informações do mapeamento entre ontologias. Por exemplo, considere as quatro ontologias construídas conforme a seção 6.1.1. Na FIG. 6.16 têm-se informações sobre o mapeamento realizado

entre a ontologia CPAD e PESSOAL e na FIG. 6.17 o grafo correspondente. Na FIG. 6.18 tem-se o mapeamento realizado entre as ontologias LEGISMAR e CPAD e na FIG. 6.19 o grafo correspondente a este mapeamento, levando em consideração os mapeamentos anteriores. Na FIG. 6.20 tem-se o mapeamento entre as ontologias PESSOAL e CPAD e na FIG. 6.21 o grafo incluindo os mapeamentos anteriores. Pode-se observar que até o presente momento não existe mapeamento envolvendo a ontologia SIGDEM, razão pelo qual, o sistema não dará início a geração da OE (FIG. 6.22). No momento em que se inclui o mapeamento da FIG. 6.23, entre as ontologias SIGDEM e LEGISMAR tem-se o grafo da FIG. 6.24 e o início da geração da OE pelo *OntoEmerge*.

Para verificação da conexidade do grafo (FIG. 6.24) levou em consideração o seu grafo subjacente, considerando as arestas paralelas como somente uma aresta.

```

...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file:/D:/MAPS/CPAD.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file:/D:/MAPS/PESSOAL.rdfs"/>
4 </a:Mapping>
...

```

FIG. 6.16 Informação de mapeamento entre as ontologias CPAD e PESSOAL

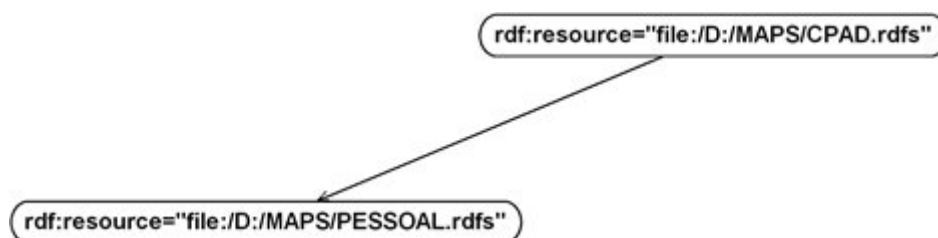


FIG. 6.17 Grafo construído com a informação de mapeamento entre as ontologias CPAD e PESSOAL

```

...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file:/D:/MAPS/LEGISMAR.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file:/D:/MAPS/CPAD.rdfs"/>
4 </a:Mapping>
...

```

FIG. 6.18 Informação de mapeamento entre as ontologias LEGISMAR e CPAD

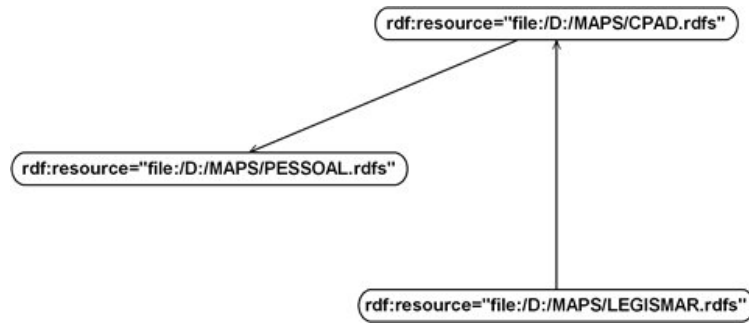


FIG. 6.19 Grafo construído com a inclusão da informação de mapeamento entre as ontologias LEGISMAR e CPAD

```

...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file:/D:/MAPS/PESSOAL.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file:/D:/MAPS/CPAD.rdfs"/>
4 </a:Mapping>
...

```

FIG. 6.20 Informação de mapeamento entre as ontologias PESSOAL e CPAD

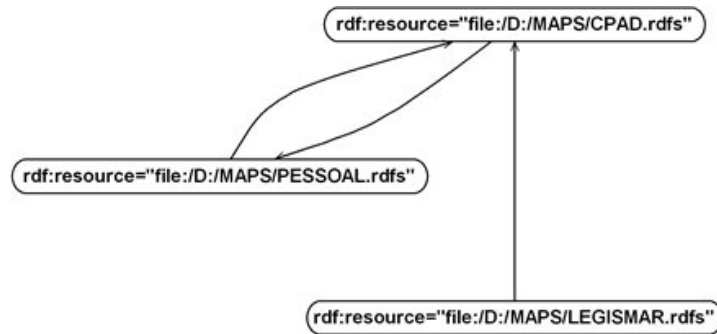


FIG. 6.21 Grafo construído com a inclusão da informação de mapeamento entre as ontologias CPAD e PESSOAL

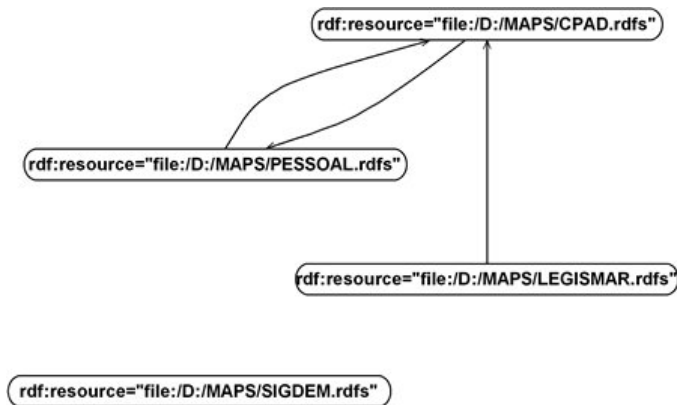


FIG. 6.22 Grafo com as informações dos mapeamentos entre ontologias

```

...
1 <a:Mapping rdf:ID="">
2   <a:relatesSourceOntology rdf:resource="file:/D:/MAPS/SIGDEM.rdfs"/>
3   <a:relatesTargetOntology rdf:resource="file:/D:/MAPS/LEGISMAR.rdfs"/>
4 </a:Mapping>
...

```

FIG. 6.23 Informação de mapeamento entre as ontologias SIGDEM e LEGISMAR

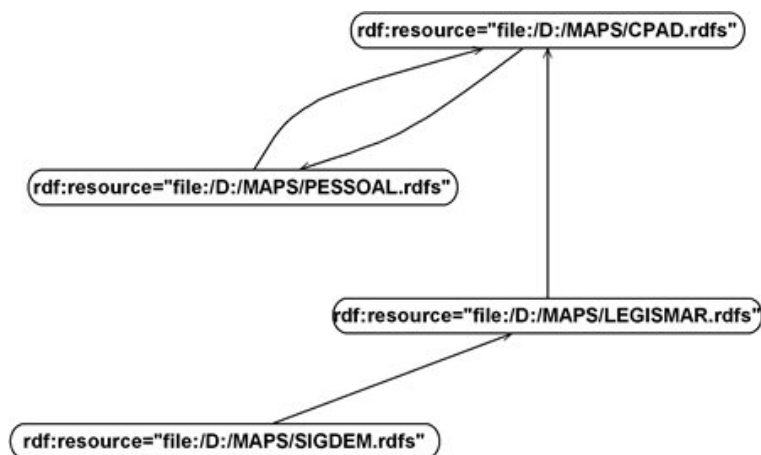


FIG. 6.24 Grafo construído com a inclusão da informação de mapeamento entre as ontologias SIGDEM e LEGISMAR

A seguir, para os mapeamentos entre conceitos apresentados nas figuras 6.25, 6.27, 6.29, 6.31, 6.33, 6.35, 6.37 e 6.39, o sistema representa estes mapeamentos através dos grafos direcionados das figuras 6.26, 6.28, 6.30, 6.32, 6.34, 6.36, 6.38 e 6.40, respectivamente.

```

...
1 <a:ConceptBridge rdf:ID="i-1193416327078-1265512569">
2   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
3   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#DOCUMENTOS"/>
4 </a:ConceptBridge>
...

```

FIG. 6.25 Exemplo de um mapeamento entre conceitos - ponte semântica (*ConceptBridge*)

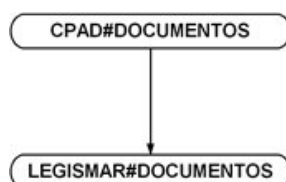


FIG. 6.26 Grafo construído com a informação dos mapeamento entre os conceitos CPAD#DOCUMENTOS e LEGISMAR#DOCUMENTOS

```

...
1 <a:ConceptBridge rdf:ID="i-1193416327078-1265512570">
2   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/CPAD#OM"/>
3   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
4 </a:ConceptBridge>
...

```

FIG. 6.27 Exemplo de um mapeamento entre conceitos - ponte semântica (*ConceptBridge*)



FIG. 6.28 Grafo construído com a inclusão da informação de mapeamento entre os conceitos CPAD#OM e LEGISMAR#OM

```

...
1 <a:ConceptBridge rdf:ID="i-1193416327078-1265512571">
2   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
3   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/PESSOAL#DOCUMENTOS"/>
4 </a:ConceptBridge>
...

```

FIG. 6.29 Exemplo de um mapeamento entre conceitos - ponte semântica (*ConceptBridge*)



FIG. 6.30 Grafo construído com a inclusão da informação de mapeamento entre os conceitos CPAD#DOCUMENTOS e PESSOAL#DOCUMENTOS

```

...
1 <a:ConceptBridge rdf:ID="i-1193416327078-1265512572">
2   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#DOCUMENTOS"/>
3   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
4 </a:ConceptBridge>
...

```

FIG. 6.31 Exemplo de um mapeamento entre conceitos - ponte semântica (*ConceptBridge*)



FIG. 6.32 Grafo construído com a inclusão da informação de mapeamento entre os conceitos LEGISMAR#DOCUMENTOS e CPAD#DOCUMENTOS

```

...
1 <a:ConceptBridge rdf:ID="i-1193416327078-1265512573">
2   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
3   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/CPAD#OM"/>
4 </a:ConceptBridge>
...

```

FIG. 6.33 Exemplo de um mapeamento entre conceitos - ponte semântica (ConceptBridge)



FIG. 6.34 Grafo construído com a inclusão da informação de mapeamento entre os conceitos LEGISMAR#OM e CPAD#OM

```

...
1 <a:ConceptBridge rdf:ID="i-1193416327078-1265512574">
2   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/PESSOAL#DOCUMENTOS"/>
3   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
4 </a:ConceptBridge>
...

```

FIG. 6.35 Exemplo de um mapeamento entre conceitos - ponte semântica (ConceptBridge)



FIG. 6.36 Grafo construído com a inclusão da informação de mapeamento entre os conceitos PESSOAL#DOCUMENTOS e CPAD#DOCUMENTOS


```

...
1 <a:ConceptBridge rdf:ID="i-1193416327078-1265512575">
2   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/SIGDEM#EXPEDIENTE"/>
3   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#DOCUMENTOS"/>
4 </a:ConceptBridge>
...

```

FIG. 6.37 Exemplo de um mapeamento entre conceitos - ponte semântica (*ConceptBridge*)

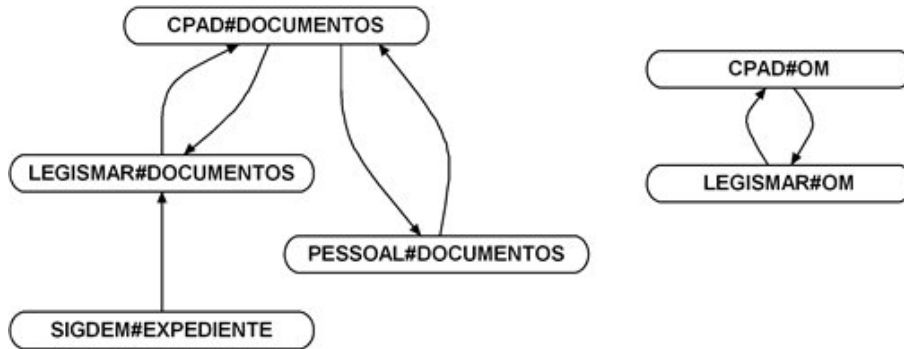


FIG. 6.38 Grafo construído com a inclusão da informação de mapeamento entre os conceitos SIGDEM#EXPEDIENTE e LEGISMAR#DOCUMENTOS

```

...
1 <a:ConceptBridge rdf:ID="i-1193416327078-1265512576">
2   <a:relatesSourceEntity rdf:resource="http://protege.stanford.edu/SIGDEM#ORGAO"/>
3   <a:relatesTargetEntity rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
4 </a:ConceptBridge>
...

```

FIG. 6.39 Exemplo de um mapeamento entre conceitos - ponte semântica (*ConceptBridge*)

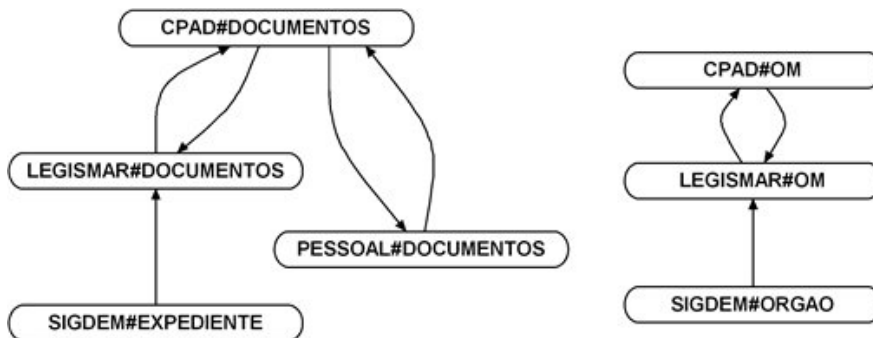


FIG. 6.40 Grafo construído com a inclusão da informação de mapeamento entre os conceitos SIGDEM#ORGAO e LEGISMAR#OM

Após a construção deste grafo direcionado, obtido com os conceitos e os mapeamentos entre conceitos, o sistema inicia a separação dos componentes conexos de seu grafo subjacente, considerando as arestas paralelas como somente uma aresta, que são os subgrafos de Conceitos Afins (definição básica 16). Na FIG. 6.40, observa-se dois subgrafos

construídos, com base nesses conceitos. Para separação, de suas componentes conexas foi utilizado a busca em profundidade no grafo subjacente construído.

O *ranking* dos conceitos (definição básica 19) de cada subgrafo (Conceitos Afins) é então calculado e em seguida os conceitos de maior *ranking* (heurística 2) serão separados dentro dos Conceitos Afins (FIG. 6.41 e FIG. 6.42), para serem representados na OE. No exemplo têm-se os conceitos CPAD#DOCUMENTOS (0.405) e LEGISMAR#OM (0.405) com os maiores rankings em cada um dos subgrafos (FIG. 6.41 e FIG. 6.42). Neste exemplo, foi considerada a pontuação igual a 1, para todos os mapeamentos entre conceitos.

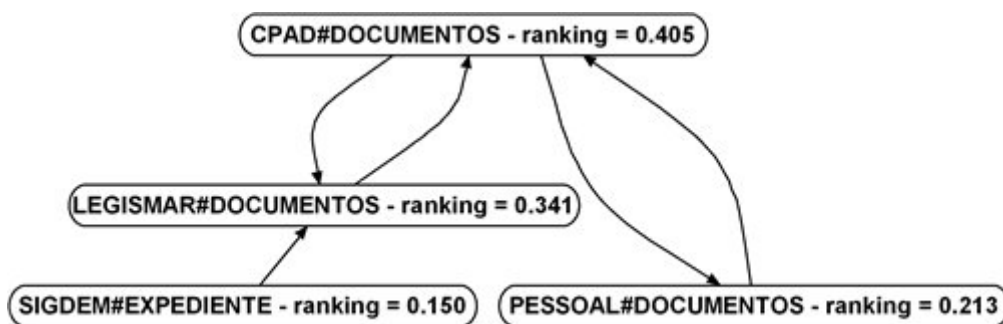


FIG. 6.41 Subgrafo com as informações de *ranking*

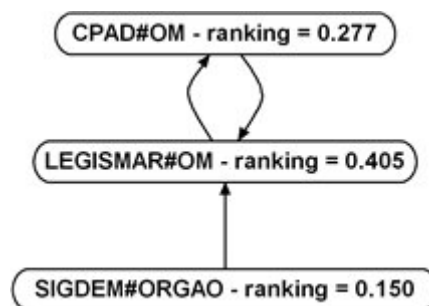


FIG. 6.42 Subgrafo com as informações de *ranking* dos conceitos

Definidos os conceitos da OE o sistema *OntoEmerge* iniciará a escolha de suas propriedades e relações com base nas heurísticas 3 e 4 do capítulo 4.

A OE gerada é apresentada graficamente na FIG. 6.43 e em RDF na FIG. 6.44, onde se observa que os conceitos eleitos foram CPAD#DOCUMENTOS e LEGISMAR#OM, em virtude da aplicação da heurística 2. Para escolha atributos e relações dos conceitos eleitos foram utilizadas as heurísticas 3 e 4 para recuperação das ontologias correspondentes aos conceitos eleitos.

Pode-se observar que o *range* das relações CPAD#DESTINO e CPAD#ORIGEM foi modificado para LEGISMAR#OM, em virtude do conceito CPAD#OM não ter sido um conceito eleito.

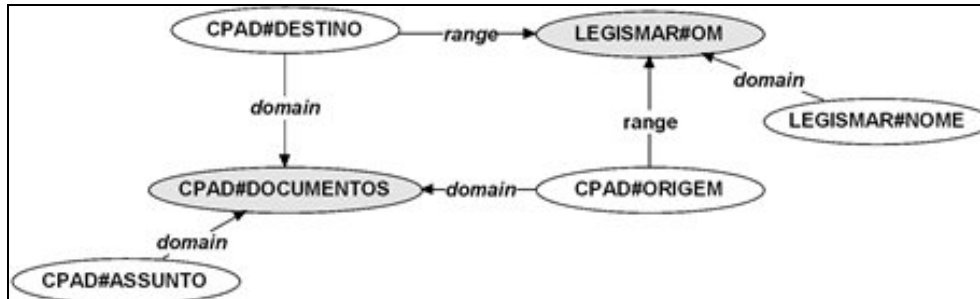


FIG. 6.43 Representação final da OE

```

<rdf:RDF
  xmlns="http://www.owl-ontologies.com/EMERGE_ONTO.owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:p1="http://protege.stanford.edu/CPAD#"
  xmlns:p2="http://protege.stanford.edu/LEGISMAR#"
  xmlns:daml="http://www.daml.org/2001/03/daml+oil#"

  <rdf:Description rdf:about="http://www.owl-ontologies.com/EMERGE_ONTO.owl">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Ontology"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/CPAD#DOCUMENTOS">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/CPAD#ASSUNTO">
    <rdfs:domain rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/CPAD#DESTINO">
    <rdfs:range rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
    <rdfs:domain rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/CPAD#ORIGEM">
    <rdfs:range rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
    <rdfs:domain rdf:resource="http://protege.stanford.edu/CPAD#DOCUMENTOS"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/LEGISMAR#OM">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://protege.stanford.edu/LEGISMAR#NOME">
    <rdfs:domain rdf:resource="http://protege.stanford.edu/LEGISMAR#OM"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
  </rdf:Description>
</rdf:RDF>

```

FIG. 6.44 Representação final da OE em RDF

6.2 AVALIAÇÃO DO PROTÓTIPO ONTOEMERGE

A fim de verificar e avaliar a conformidade dos objetivos definidos para o sistema proposto neste trabalho com as funcionalidades atingidas na implementação foram configurados alguns testes com uma base maior de ontologias e mapeamentos. Estes testes visam verificar se a OE gerada retrata uma visão global significativa dos mapeamentos pré-estabelecidos e disponíveis nos diversos pontos.

Como observado anteriormente, o sistema inicia a geração da OE quanto atinge uma determinada quantidade de mapeamentos. Assim, o planejamento dos testes aqui propostos consiste em desenvolver vários conjuntos de documentos de mapeamento, que serão utilizados pelo *OntoEmerge*, cujas OEs geradas constituirão a base dessa análise. Para a geração destes mapeamentos foram utilizadas as ontologias e ferramentas disponibilizadas pelo *Benchmark* proposto pela *Ontology Alignment Evaluation Initiative – OAEI*.

Para realização desses experimentos foi desenvolvida uma bancada de testes cuja descrição e análises serão feitas nas próximas seções.

6.2.1 DESCRIÇÃO DAS ONTOLOGIAS UTILIZADAS PELA BANCADA DE TESTES

Desde 2004 a *OAEI* vem organizando eventos com o objetivo de avaliar as tecnologias de casamento (*matching*) de ontologias. Em 2007 foi produzido um *benchmark* que visa identificar os pontos fortes e fracos de cada algoritmo de casamento. Este *benchmark* possui 51 ontologias, uma ontologia do domínio da área de bibliografia (chamada de ontologia de referência - OR) e outras ontologias alternativas para realização de casamentos e posterior avaliação dos algoritmos.

A ontologia de referência é baseada em uma visão subjetiva de como uma ontologia da área bibliográfica deveria ser e em quatro ontologias que foram objeto de experimentos de projetos e de uso, e que serão também utilizadas nos testes. A ontologia de referência possui 33 classes nomeadas, 24 propriedades (relacionamentos), 40 propriedades (atributos), 56 *named individuals* e 20 *anonymous individuals*.

Os testes são gerados sistematicamente entre a ontologia de referência e demais ontologias alternativas, com o objetivo de devolver um número de informações que permita avaliar a forma como o algoritmo de casamento se comporta.

As ontologias alternativas possuem conceitos que retratam seis tipos de alterações em relação à ontologia de referência, que são:

- Nome: os nomes das entidades podem ser nomes gerados randomicamente, sinônimos, nomes com diferentes convenções e *strings* em outro idioma, diferente da inglesa;
- Comentário: comentários podem ser suprimidos ou traduzidos para outra linguagem;
- Especialização da hierarquia: poderá ser suprimida, expandida ou diminuída;
- Instâncias: podem ser suprimidas;
- Propriedades: podem ser suprimidas ou ter as restrições nas classes desconsideradas; e
- Classes: suas relações com outras classes podem ser expandidas ou diminuídas.

Foi verificado que algumas ontologias alternativas possuem informações que não pertencem ao domínio da área bibliográfica. Por exemplo, em um dos testes propostos, a ontologia de referência é casada com uma ontologia alternativa que possui os nomes de seus conceitos gerados de forma randômica. Sendo assim, não geram informações de casamentos entre as mesmas, que indique os conceitos similares.

Devido à necessidade de utilização destas ontologias pela bancada de testes proposta nesta dissertação, a escolha das mesmas foi realizada levando em consideração aquelas que tenham possibilidade de recuperar e apontar entidades similares.

Foram analisados os diversos testes realizados pela *OEAI* com o intuito de selecionar as ontologias que serão utilizadas pela bancada de teste, aqui apresentada. Estes testes são descritos em seguida, sendo que aqueles cuja ontologia foi escolhida possuem após seu nome uma identificação numérica entre parênteses (ex. 101 (0) - "101" é o nome do teste e "0" é a identificação numérica dada a ontologia do teste, que indica também que a mesma foi selecionada).

- 101 (0) - Teste de Conceito. Comparação da ontologia de referência com ela mesma. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 102 - Teste de Conceito. Comparação da ontologia de referência com outra totalmente diferente, sendo que a utilizada foi a ontologia *Food*. A ontologia deste teste não será utilizada pela bancada de teste proposta nesta dissertação, devido ontologia *Food* não ser da área Bibliográfica.

- 103 (1) - Teste de Conceito: generalização da linguagem. Compara a ontologia de referência com sua generalização em *OWL Lite* (isto é, restrições que não estão disponíveis são substituídas por outras mais genéricas existentes). A generalização basicamente remove as *owl:unionOf*, *owl:oneOf* e as *Property types* (*owl:TransitiveProperty*). Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 104 (2) - Teste de Conceito: restrição da linguagem. Compara a ontologia de referência com sua restrição em *OWL Lite* (onde as restrições que não estão disponíveis são descartadas). Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 201 - Sem nomes. Todo rótulo ou identificador é substituído por outro gerado aleatoriamente. A ontologia deste teste não será utilizada pela bancada de testes, proposta nesta dissertação, devido os rótulos e identificadores da ontologia gerados aleatoriamente não retratarem uma ontologia da área bibliográfica.
- 202 - Sem nomes. Todo rótulo ou identificador é substituído por outro gerado aleatoriamente, sendo que os comentários são excluídos. A ontologia deste teste não será utilizada pela bancada de testes proposta nesta dissertação, devido os rótulos e identificadores da ontologia gerados aleatoriamente não retratarem uma ontologia da área bibliográfica.
- 203 - Sistemático: palavras escritas de forma errada. Erros de ortografia são gerados aleatoriamente e aplicados aos rótulos e comentários. Ficaram dúvidas quanto à aplicação destes erros de ortografia nos comentários, em virtude da tabela que resume os testes dizer que os comentários serão suprimidos. Os resultados deste teste não foram disponibilizados pela *OAEI*. A ontologia deste teste não será utilizada pela bancada de teste proposta nesta dissertação, devido à ontologia não estar disponível.
- 204 (3) - Sistemático: convenção de nomes. Diferentes convenções de nomes são aplicadas aos rótulos (letras maiúsculas e/ou minúsculas uso de sublinhado para separar palavras, traços, etc). Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 205 (4) - Sistemático: sinônimos. Os rótulos são substituídos por sinônimos. Os comentários são omitidos. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).

- 206 (5) - Sistemático: nomes estrangeiros. A ontologia completa é traduzida para outro idioma (rótulos, identificados e comentários). Neste teste foi realizada a tradução de inglês para francês. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 207 (6) - Sistemático. Os rótulos e identificadores são trazidos para outro idioma (neste teste foi realizada a tradução de inglês para francês). Os comentários não são traduzidos. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 208 (7) - Sistemático. Diferentes convenções de nomes são utilizadas, tais como: nomes escritos com letras minúsculas, no plural, com underscore e abreviados. Os comentários são suprimidos. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 209 (8) - Sistemático. Nomes das entidades são alterados por sinônimos. Os comentários são omitidos. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 210 (9) - Sistemático. A ontologia completa é traduzida para outro idioma (rótulos, identificadores). Neste teste foi realizada a tradução de inglês para francês. Os comentários são omitidos. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 221 (10) - Sistemático: sem hierarquia. Todas as declarações de subclasses de classes nomeadas são omitidas. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 222 (11) - Sistemático: com a hierarquia reduzida. A hierarquia existe, mas de forma reduzida. Possui 28 classes, 24 propriedades (relacionamentos) e 38 propriedades (atributos).
- 223 (12) - Sistemático: com a hierarquia expandida. Algumas classes são criadas e colocadas como subclasses de outras, aumentando assim o número de classes. Possui 66 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 224 (13) - Sistemático: sem instâncias. Todas as instâncias são omitidas da ontologia. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 225 (14) - Sistemático: sem restrições. Todas as restrições locais das propriedades são omitidas. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).

- 226 - Sistemático: sem propriedades (atributos). Todos os propriedades (atributos) são convertidos para xsd:string. Os resultados deste teste não foram disponibilizados pela OAEI.
- 227 - Sistemático: diferenças de unidades. Valores (limitados) são expressos em diferentes propriedades (atributos). Os resultados deste teste não foram disponibilizados pela OAEI.
- 228 (15) - Sistemático: sem propriedades (atributos) e propriedades (relacionamentos). As propriedades (atributos) e propriedades (relacionamentos) são suprimidas. Possui 33 classes, 0 propriedade (relacionamento) e 0 propriedade (atributo).
- 229 - Sistemático: classes versus instâncias. Algumas classes tornaram-se instâncias. Os resultados deste teste não foram disponibilizados pela OAEI. A ontologia deste teste não será utilizada pela bancada de testes proposta nesta dissertação, devido à ontologia não estar disponível.
- 230 (16) - Sistemático: redução das entidades. Alguns componentes das classes foram expandidos dentro da estrutura da classe (por exemplo: os atributos ano, mês e dia no lugar de data). Ex: A ontologia de referência possui uma classe chamada Date, mas na ontologia deste teste a classe Date é suprimida e são utilizados os atributos ano, mês e dia como propriedades de outra classe. Possui 25 classes, 8 propriedades (relacionamentos) e 42 propriedades (atributos).
- 231 - Sistemático: aumento das entidades. Os resultados deste teste não foram disponibilizados pela OAEI. A ontologia deste teste não será utilizada pela bancada de testes proposta nesta dissertação, devido à ontologia não estar disponível.
- 232 (17) - Sistemático. São retiradas as hierarquias, isto é, todas as sub-classes e as instâncias são suprimidas. Possui 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos).
- 233 (18) - Sistemático. As hierarquias são retiradas, isto é, todas as sub-classes e as propriedades (relacionamentos) e propriedades (atributos) correspondentes. Possui 33 classes, 0 propriedade (relacionamento) e 0 propriedade (atributo).
- 236 (19) - Sistemático. Foram retiradas as propriedades (relacionamentos), propriedades (atributos) e as instâncias. Possui 33 classes, 0 propriedade (relacionamento) e 0 propriedade (atributo).
- 237 (20) - Sistemático. A hierarquia foi reduzida e as instâncias retiradas. Possui 28 classes, 24 propriedades (atributos) e 40 propriedades (relacionamentos).

- 238 (21) - Sistemático. A hierarquia foi expandida e as instâncias retiradas. Possui 66 classes, 24 propriedades (relacionamentos) e 39 propriedades (atributos).
- 239 (22) - Sistemático. A hierarquia foi reduzida e as propriedades (atributos) e as propriedades (relacionamentos) retiradas. Possui 28 classes, 0 propriedade (relacionamento) e 0 propriedade (atributo).
- 240 (23) - Sistemático. A hierarquia foi expandida e as propriedades (atributos) e as propriedades (relacionamentos) retiradas. Possui 66 classes, 0 propriedade (relacionamento) e 0 propriedade (atributo).
- 241 (24) - Sistemático. A hierarquia retirada, isto é, todas as sub-classes e instâncias e os propriedades (relacionamentos) e propriedades (atributos) são suprimidos. Possui 33 classes, 0 propriedade (relacionamento) e 0 propriedade (atributo).
- 246 (25) - Sistemático. A hierarquia foi reduzida e as instâncias e as propriedades (atributos) e propriedades (relacionamentos) foram suprimidos. Houve a diminuição de 5 classes, em relação à ontologia de referência. Possui 28 classes, 0 propriedade (relacionamento) e 0 propriedade (atributo).
- 247 (26) - Sistemático. Hierarquia expandida e as instâncias e as propriedades (atributos) e propriedades (relacionamento) são suprimidos. Possui 66 classes, 0 propriedade (relacionamento) e 0 propriedade (atributo).
- 248 a 254, 257 a 262 e 265 e 266 - Sistemático. Os nomes de suas entidades são gerados randomicamente, sendo que não se ajustam ao domínio da área bibliográfica. Exemplo de um nome de classe gerado randomicamente: a362042121.
- 301 (27) - Ontologia real: BibTeX/MIT. Comparação da ontologia de referência com uma ontologia real, mais simples e próxima da ontologia BibTeX. Possui 14 classes, 1 propriedade (relacionamento) e 40 propriedades (atributos).
- 302 (28) - Ontologia real: BibTeX/UMBC. Comparação da ontologia de referência com uma ontologia similar a anterior, sendo mais próxima a ontologia BibTeX, com diferentes convenções de extensões e nomes. Possui 12 classes, 4 propriedades (relacionamentos) e 24 propriedades (atributos).
- 303 (29) - Ontologia real: Karlsruhe. Comparação da ontologia de referência com uma ontologia real, que contém mais itens que a ontologia de referência. Possui 56 classes, 72 propriedades (relacionamentos) e 0 propriedade (atributo).

- 304 (30) - Ontologia real: INRIA. Este teste compara a ontologia de referência com uma ontologia real, muito próxima da ontologia BibTeX. Possui 39 classes, 37 propriedades (relacionamentos) e 11 propriedades (atributo).

Após a análise dos testes descritos anteriormente, foram selecionadas 31 ontologias para compor a bancada de testes. As ontologias que se utilizam de mecanismos para geração randômica de termos não foram utilizadas, já que os casamentos não geram termos compatíveis ao contexto da área bibliográfica. Conforme, dito anteriormente as ontologias selecionadas de cada teste possui uma identificação entre parênteses (0 a 30) após o nome do teste.

6.2.2 DESCRIÇÃO DA BANCADA DE TESTES

Os testes foram feitos com a finalidade de validar a representatividade da OE, levando em consideração alguns dados quantitativos computados dos arquivos de mapeamento e da própria OE.

No trabalho (GANGEMI et al, 2006) propõem-se medidas qualitativas (estruturais, funcionais e usabilidade-perfil) para avaliação das ontologias, porém não está no escopo deste trabalho avaliar qualitativamente as ontologias geradas. Desta forma, foram identificadas algumas medidas quantitativas para avaliarmos preliminarmente a proposta.

As implementações e abordagens utilizadas são apresentadas na FIG. 6.45 e descritas em seguida.

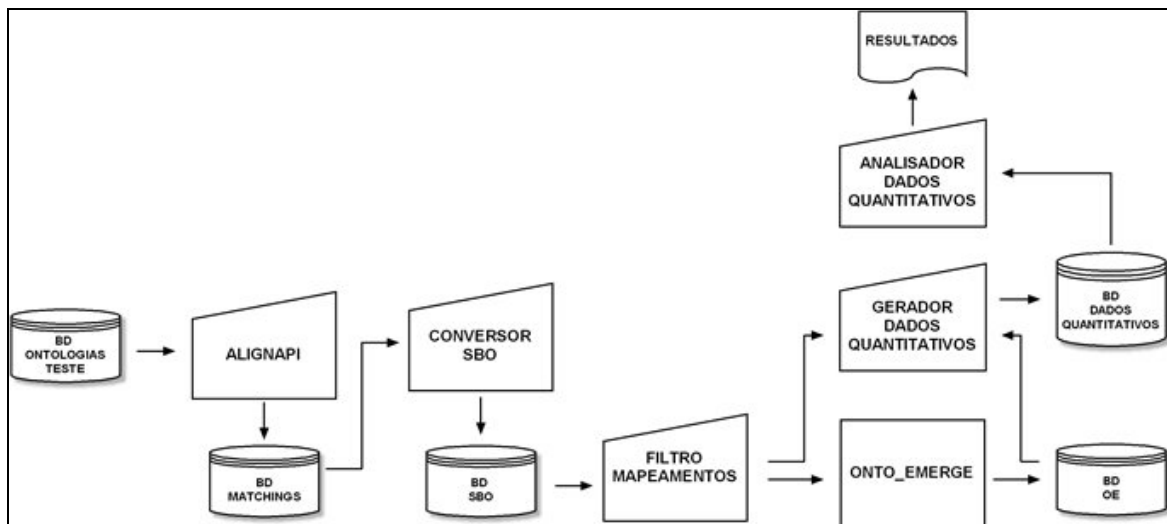


FIG. 6.45 Arquitetura da Bancada de teste

- **Alignapi** - o componente *Alignapi* (EUZENAT, 2004) é uma API apresentada pela OAEI e tem como finalidade estabelecer *matchings* entre conceitos de ontologias.

Primeiramente a bancada de testes gera casamentos entre os conceitos envolvendo as 31 ontologias gerando 930 arquivos com casamentos utilizando o *Alignapi*, isto é, para cada ontologia é realizado 30 casamentos, totalizando 930. As 31 ontologias utilizadas são armazenadas no BD ONTOLOGIAS TESTE e os arquivos com os casamentos são armazenados no BD MATCHING.

Os casamentos obtidos são gerados em formato XML conforme o exemplo apresentado na FIG. 6.46. Nas linhas 14 e 15 são indicadas as ontologias envolvidas no processo de casamento de seus conceitos e nas linhas 17 a 21 apresentam os conceitos que foram associados. Neste arquivo também é registrada a medida de similaridade atribuída ao casamento dos conceitos (linhas 21 e 29). No exemplo da FIG. 6.46 a medida é igual a 1, devido aos conceitos serem representados por termos iguais.

```

1 <?xml version='1.0' encoding='utf-8' standalone='no'?>
2 <rdf:RDF xmlns='http://knowledgeweb.semanticweb.org/heterogeneity/alignment'
3 xml:base='http://knowledgeweb.semanticweb.org/heterogeneity/alignment'
4 xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
5 xmlns:xsd='http://www.w3.org/2001/XMLSchema#'>
6 <Alignment>
7 <xml>yes</xml>
8 <level>0</level>
9 <type>*</type>
10 <time>31</time>
11 <method>fr.inrialpes.exmo.align.impl.method.StringDistAlignment</method>
12 <onto1>file:///C:/ALIGN_ONTO/onto0.rdf</onto1>
13 <onto2>file:///C:/ALIGN_ONTO/onto1.rdf</onto2>
14 <uri1>file:///C:/ALIGN_ONTO/onto0.rdf</uri1>
15 <uri2>file:///C:/ALIGN_ONTO/onto1.rdf</uri2>
16 <map>
17 <Cell>
18 <entity1 rdf:resource='http://oaei.ontologymatching.org/2007/benchmarks/101/onto.rdf#Monograph' />
19 <entity2 rdf:resource='http://oaei.ontologymatching.org/2007/benchmarks/103/onto.rdf#Monograph' />
20 <relation>=</relation>
21 <measure rdf:datatype='http://www.w3.org/2001/XMLSchema#float'>1.0</measure>
22 </Cell>
23 </map>
24 <map>
25 <Cell>
26 <entity1 rdf:resource='http://oaei.ontologymatching.org/2007/benchmarks/101/onto.rdf#Unpublished' />
27 <entity2 rdf:resource='http://oaei.ontologymatching.org/2007/benchmarks/103/onto.rdf#Unpublished' />
28 <relation>=</relation>
29 <measure rdf:datatype='http://www.w3.org/2001/XMLSchema#float'>1.0</measure>
30 </Cell>
31 </map>
32 </Alignment>
33 </rdf:RDF>

```

FIG. 6.46 Representação XML da combinação entre ontologias

- **Conversor SBO estendido** - o componente *Conversor SBO estendido* desenvolvido tem como finalidade transformar os arquivos de casamentos, gerados em formato XML próprio do mecanismo *Alignapi*, no formato SBO estendido. Apesar do *Alignapi* possibilitar que os casamentos sejam exportados para nove diferentes formatos, foi necessário desenvolver um conversor para formato *SBO estendido*, necessário para este trabalho.
- **Filtro de mapeamentos** - este componente seleciona aleatoriamente uma determinada quantidade de mapeamentos do BD SBO.
- **Gerador de dados quantitativos** - este componente quantifica algumas informações dos dados de teste e da OE gerada pelo protótipo *OntoEmerge*, armazenando-as no BD DADOS QUANTITATIVOS. Estas quantificações são:
 - Quantidade de conceitos da OE (QtdcOE);
 - Quantidade de propriedades da OE (QtdpOE); e
 - Quantidade de mapeamentos utilizados para geração da OE (Qtdmap);
- **Analizador dados quantitativos** - compara os dados disponíveis no BD DADOS QUANTITATIVOS e gera os resultados em arquivos texto, os quais podem ser visualizados por ferramentas gráficas.

6.2.3 EXPERIMENTOS REALIZADOS

Com base nas ontologias selecionadas, a bancada de testes gerou 930 arquivos de mapeamentos, que é o resultado do mapeamento entre pares de ontologias, os quais foram utilizados nos experimentos, estabelecidos na tabela 1 e descritos em seguida:

- *experimento 1* - são utilizadas as 31 ontologias. Sabe-se que o total de mapeamentos envolvendo todas estas ontologias é de 930, então foram geradas 1 OE envolvendo os 930 mapeamentos e 20 OEs para cada 100, 200, 500 e 700 mapeamentos selecionados aleatoriamente destes 930 mapeamentos e, em seguida, realizada a média dos conceitos e propriedades das 20 OEs geradas por estes 100, 200, 500 e 700 mapeamentos;
- *experimento 2* - foram selecionadas as ontologias 0-11, 13, 14, 17, que possuem 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos), 16 e 20 que possuem 28 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos). Esta seleção de ontologias foi feita com o objetivo de eliminar do conjunto de testes as ontologias sem propriedades (atributos e relacionamentos) e manter ontologias com aproximadamente o mesmo porte, ou seja, com um número de classes e propriedades similar. Os mapeamentos envolvendo todas estas ontologias somam 272. Foi gerada 1 OE envolvendo os 272 mapeamentos e 20 OEs para cada 50, 100 e 200 mapeamentos selecionados aleatoriamente destes 272 e, em seguida, foi realizada a média dos conceitos e propriedades das 20 OEs geradas por estes 50, 100 e 200 mapeamentos;
- *experimento 3* - foram selecionadas as ontologias 0-11, 13, 14 e 17, sendo que possuem 33 classes, 24 propriedades (relacionamentos) e 40 propriedades (atributos) e as quatro ontologias reais (27, 28, 29 e 30). Esta seleção de ontologias foi feita para incorporar ao conjunto de ontologias do experimento 2, as ontologias reais. Os mapeamentos envolvendo todas estas ontologias somam 342. Foi gerada 1 OE envolvendo os 342 mapeamentos e 20 OEs para cada 50, 100, 200 e 300 mapeamentos selecionados aleatoriamente destes 342 mapeamentos e, em seguida, realizada a média dos conceitos e propriedades das 20 OEs geradas por estes 50, 100, 200, e 300 mapeamentos;

A escolha dos subconjuntos de mapeamentos foi aleatória, descartando-se os casos em que o subconjunto não atendia à heurística 2, isto é, não envolvia todos os nós.

	Qtd de ontologias	Qtd total de mapeamentos envolvendo todas as ontologias utilizadas	Qtd de mapeamentos selecionados	Qtd OEs geradas
1	31	930	930	1
			700	20
			500	20
			200	20
			100	20
2	17	272	272	1
			200	20
			100	20
			50	20
3	19	342	342	1
			200	20
			100	20
			50	20

TAB. 6.1 Experimentos realizados

6.2.4 ANÁLISE DOS RESULTADOS OBTIDOS

Experimento 1 - os resultados são apresentados no gráfico da FIG. 6.47a e FIG. 6.47b. Pode-se observar, na FIG. 6.47a, que a OE gerada para 930 mapeamentos possui 123 conceitos e 85 propriedades, e será considerada como referência para análise das demais. Este resultado apresenta a quantidade máxima de conceitos que poderiam ser representados em uma OE em virtude de envolverem todos os mapeamentos possíveis entre todas as ontologias. Nos outros resultados a média dos conceitos das OEs geradas se aproxima do resultado de 930 mapeamentos. Sendo assim não existe a necessidade de que todos os pontos realizem todos os mapeamentos possíveis para que se tenha uma OE quantitativamente representativa. No

resultado de 100 mapeamentos a quantidade de conceitos ficou um pouco abaixo de 100, em virtude de possuir menor quantidade de mapeamentos entre as ontologias.

No resultado de 930 mapeamentos, FIG. 6.47b, a quantidade de propriedades ficou abaixo de 100, sendo que nos resultados de 200 e 100 mapeamentos ficaram abaixo e próximos ao resultado de 930. Nos resultados de 700 e 500 mapeamentos ficaram acima de 100 e acima e próximos ao resultado de 930. No caso das propriedades, devido à escolha dos mapeamentos ser aleatória, pode resultar em mapeamentos que contenham conceitos com mais propriedades. Por exemplo, têm-se dois conceitos Afins A e B, com 7 e 10 propriedades respectivamente. Se o conceito A foi eleito na OE do resultado de 930 mapeamentos, no resultado de 700 mapeamentos pode acontecer o caso do conceito A não entrar em nenhum mapeamento, e o conceito eleito ser o conceito B, que neste caso tem uma quantidade maior de propriedades. Tem-se assim, o motivo da diferença da quantidade de propriedades não seguir regularmente um padrão de crescimento, semelhante ao comportamento dos conceitos. Note que A e B para o resultado de 930 mapeamentos seriam conceitos Afins, mas para o resultado de 700 mapeamentos, o conceito A provavelmente não faz parte de nenhum conjunto de conceitos afins.

Embora haja diferenças das quantidades de propriedades em relação ao resultado de 930 mapeamentos e os demais, a escolha das propriedades são aquelas dos conceitos eleitos, desta forma a estratégia usada pela abordagem de Ontologias Emergentes mantém-se coerente com estes resultados.

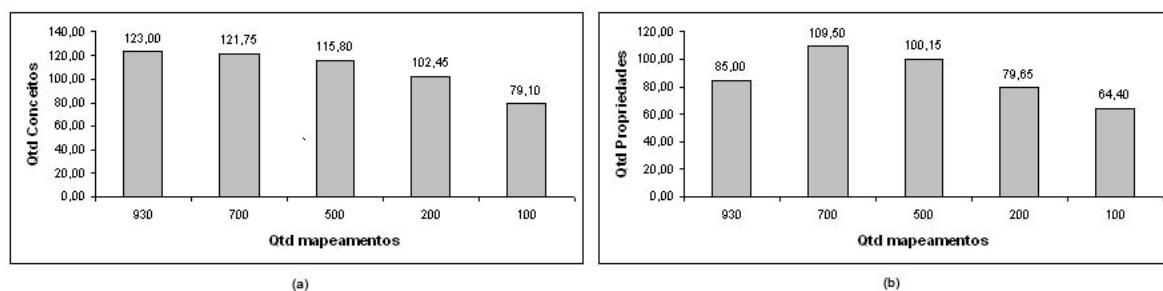


FIG. 6.47 Resultados do experimento 1

Experimento 2 - os resultados são apresentados no gráfico da FIG. 6.48a e FIG. 6.48b. A análise inicial do experimento 2 é análoga ao experimento 1, sendo que neste experimento foram selecionadas 17 ontologias as quais possuem uma média de 33 conceitos e 64

propriedades. Pelos resultados, obtidos nota-se que as OEs geradas possuem uma quantidade de conceitos superior a 33, como pode ser visualizado no gráfico da FIG. 6.48a.

Conforme reportado anteriormente, as ontologias construídas pela *OEAI* sofreram modificações em relação a ontologia de referência (e.g. sinônimos, nomes em línguas diferentes e outros). Devido a ferramenta *Alignapi* não utilizar conhecimento adicional, como tesouros, vocabulários, dicionários e outros, para resolver estes conflitos, as ontologias geradas pela bancada de testes possuem quantidades de classes superiores a 33. Por exemplo, sejam duas classes A e B (equivalentes) de ontologias distintas, sendo que a classe A é escrita na língua inglesa e a classe B em língua diferente, então estas duas classes foram representadas na OE.

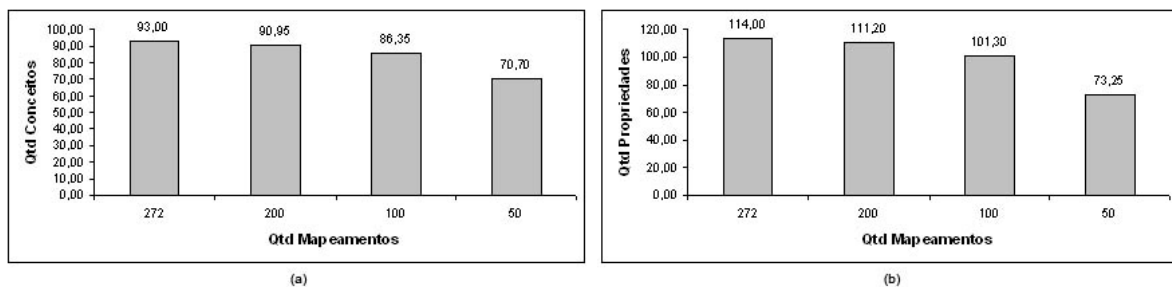


FIG. 6.48 Resultados do experimento 2

Experimento 3 - os resultados são apresentados no gráfico da FIG. 6.49a e FIG. 6.49b. A análise do experimento 3 também é análoga ao experimento 1 e 2.

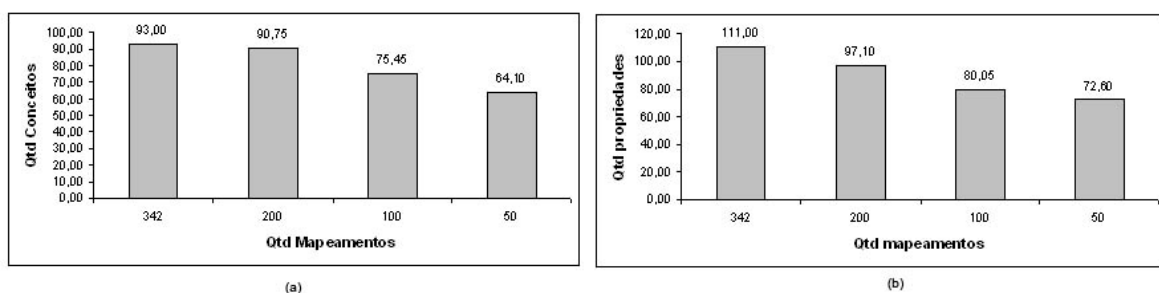


FIG. 6.49 Resultados do experimento 3

6.3 CONSIDERAÇÕES FINAIS

As quatro ontologias que serviram de base para o exemplo de uso apresentado na seção 6.1 não foram desenvolvidas por especialistas, mas serviram de base para demonstrar o protótipo e gerar uma visão global inicial. Entretanto, a idéia de tomar os sistemas da Marinha como base para a construção de ontologias setoriais parece ser um ótimo caminho para uma avaliação mais completa. Neste sentido, seria interessante iniciar um estudo de caso real, onde o objetivo seria abranger maior número de sistemas da Marinha e desenvolver ontologias específicas com base nos mesmos, e os possíveis mapeamentos, para mais tarde aplicar o protótipo *OntoEmerge* e fazer uma avaliação da OE gerada, junto aos usuários de todos os sistemas.

Os testes da seção 6.2 foram realizados somente para avaliar quantitativamente a OE e como foi observado nos experimentos a ajuda de conhecimento adicional, como tesouros e dicionários são imprescindíveis para resolver conflitos de língua e de semântica.

O *PageRank* (BRIN & PAGE, 1998) pressupõe iterações para que haja convergência para o maior valor de *Ranking*, mas no exemplo de uso utilizado (seção 6.1), estas iterações não foram realizadas devido a utilização de poucos conceitos e nos testes (seção 6.2), não foi utilizado em virtude da avaliação ser somente quantitativa, isto é, independe do conceito que foi eleito, e for facilitar o funcionamento da bancada de testes, diminuindo o tempo para realização dos experimentos.

A ferramenta *Alignapi* realizou os casamentos e conseqüentemente os mapeamentos para realização dos testes. Outras ferramentas para análise estrutural das ontologias poderiam melhorar e aumentar quantidade e qualidade dos casamentos. Além disso os mapeamentos feitos pela ferramenta não passaram pela supervisão humana, o que pode ter influenciado desfavoravelmente os resultados obtidos.

No trabalho de (LACHTIM, 2008) foi desenvolvido uma extensão da ferramenta *CROSI Mapping System (CMS)* (CMS) denominada de CMS estendido para realização e melhoria dos casamentos. O CMS estendido utiliza vários métodos de casamento, como por exemplo: *EditDistance*. Utiliza também as informações de restrições, domínio e *range* das propriedades, bem com a hierarquia das classes, para aumentar a precisão e revocação dos casamentos.

A análise qualitativa das ontologias geradas seria necessária, mas não foi possível incluir no escopo deste trabalho, pelo fato de envolver um trabalho de longa duração e alta complexidade. Em trabalhos futuros, seria necessário considerarmos o mapeamento entre as hierarquias, para avaliar qualitativamente a OE gerada pela abordagem proposta neste trabalho.

7 CONCLUSÃO

Este trabalho teve como proposta inicial a geração de uma visão global integrada a partir das informações de mapeamentos realizados entre os diversos pontos de uma organização. Partiu-se da premissa de que esses pontos organizam suas informações e conhecimento através de ontologias locais e realizam mapeamentos ponto-a-ponto quando desejam trocar informações com outros pontos.

Inicialmente, para verificar se existia alguma abordagem dessa natureza, foi realizado um estudo sobre o estado da arte em interoperabilidade de ontologias (LACHTIM et al., 2008). Neste contexto, foram encontradas quatro abordagens principais para realizar esse processo: mapeamento, alinhamento, combinação e integração. Também foram investigadas algumas ferramentas disponíveis na literatura para realizar a interoperabilidade de ontologias, em cada uma das abordagens, incluindo também ferramentas para o cálculo de similaridade a exemplo do MAFRA (MAEDCHE et al. 2002) e (NUNO SILVA, 2004) e CMS (CMS), respectivamente.

Dentre as tecnologias de integração não foi encontrada nenhuma ferramenta capaz de gerar uma visão global integrada de várias ontologias de forma automática, levando em consideração os mapeamentos realizados entre ontologias participantes dos mapeamentos.

Adicionalmente, como a proposta era integrar ontologias de diversos pontos ou sítios, as tecnologias de distribuição e integração de dados também foram investigadas, com destaque para a tecnologia P2P (BRITO, 2005).

Com os estudos realizados, uma arquitetura que denominou-se Ontologias Emegentes, foi definida com cinco módulos e três bases de dados, conforme descrito no cap.4. Com base nesta arquitetura, foi desenvolvido um protótipo, o *OntoEmerge*, através do qual foi possível gerar uma visão global das ontologias disponíveis em vários pontos. Este protótipo levou em consideração a ferramenta MAFRA para realização de mapeamentos e o COPPEER (MIRANDA, 2006) para dar suporte à transferência de instâncias a partir dos mapeamentos e ontologias mapeadas, bem como outras funcionalidades previstas nos sistemas P2P.

Para eleição dos conceitos e propriedades representados na visão global, foram especificadas heurísticas para definir os elementos envolvidos nos mapeamentos, permitindo uma melhor formalização do processo de interoperabilidade.

Para exemplificar o uso do protótipo foram definidas quatro ontologias da área de gerência de documentos, três delas com base em sistemas utilizados na Marinha e uma com base em conhecimentos subjetivos de organização de documentos.

A validação do protótipo levou em consideração 31 ontologias da área bibliográfica, sendo realizados experimentos, com resultados considerados satisfatórios.

Este trabalho foi publicado e apresentado VI *Workshop* de Teses e Dissertações em Bancos de Dados (WTDBD-2007) durante sua realização, onde foi possível, através da interação de outros acadêmicos, obter informações construtivas e, também, sua avaliação.

7.1 CONTRIBUIÇÕES

Portanto, o trabalho proposto demonstrou que é possível utilizar mapeamentos para a geração de uma visão global entre ontologias, levando a algumas contribuições importantes:

- extensão de uma arquitetura P2P para a realização de interoperação de ontologias;
- definição de heurísticas necessárias para gerar a visão global;
- estudo abrangente sobre interoperabilidade de ontologias e formalizado através do relatório técnico (LACHTIM et al, 2008);
- concretização da proposta através do desenvolvimento de um protótipo, capaz de gerar automaticamente uma visão global de ontologias distribuídas numa arquitetura P2P;
- desenvolvimento de uma bancada de testes para validação e avaliação do protótipo;
- uma abordagem inovadora para a integração de ontologias, considerando diversas ontologias de domínio ou aplicação, para facilitar a criação de uma ontologia mais abrangente.
- Refinamento da heurística 2 através da adaptação da proposta de (BRIN & PAGE, 1998).

7.2 MELHORIAS E TRABALHOS FUTUROS

Nesta seção serão apresentadas algumas sugestões de melhoria da proposta, isto é a arquitetura, e para o protótipo desenvolvido OntoEmerge. Finalizando com algumas sugestões de trabalhos futuros.

7.2.1 MELHORIAS NA PROPOSTA

Apesar do trabalho realizado ter produzido resultados satisfatórios, algumas considerações importantes merecem destaque, visando seu aprimoramento:

- nesse trabalho somente foram consideradas mapeamentos do tipo 1:1. Seria interessante estendê-lo para considerar outros tipos de mapeamentos, como por exemplo: conceito A da ontologia X mapeado para os conceitos B e C da ontologia Y;
- utilização do mecanismo de inferência aplicado aos mapeamentos, visando tornar a visão global mais robusta. Dessa forma, outros mapeamentos poderiam ser inferidos automaticamente, sem intervenção do usuário;
- levar em consideração as instâncias das ontologias para a validação dos conceitos eleitos;
- a partir dos mapeamentos das propriedades dos conceitos, outras propriedades mais importantes também poderiam ser eleitas. Atualmente, uma vez eleito, o conceito herda todas as suas propriedades;
- melhoria da qualidade semântica no processo de casamento, por meio da utilização de um conhecimento auxiliar, a exemplo de tesouros, dicionários e outros;
- geração da visão local de forma automática, tendo como referência os mapeamentos realizados pelo ponto e a OE.

7.2.2 MELHORIAS NO PROTÓTIPO ONTOMERGE

Como melhoria do protótipo pode-se citar:

- desenvolvimento de funcionalidades para geração da visão local;
- construção de interfaces gráficas, próprias do sistema, para visualização das visões local e global, bem como suas instâncias; e
- desenvolvimento do módulo de rastreamento visando monitorar algumas atividades como por exemplo a situação atual da OE.

7.2.3 SUGESTÕES PARA TRABALHOS FUTUROS

Como trabalhos futuros sugere-se o seguinte:

- adaptar a proposta aos sistemas de integração já existentes, no que tange ao processamento de consultas. Sabe-se que a troca de informações em diversos sistemas, se dá pela reescrita de consulta. O acúmulo destas reescritas poderão dar subsídios para a construção de uma visão global de forma dinâmica. O que, de modo crescente, facilitaria outras consultas e entradas de novos pontos na rede;
- adaptar a proposta para geração de uma visão integrada de ontologias em mecanismos de rastreamento, como Swoogle (SWOOGLE). Sabe-se que estes mecanismos de rastreamento proporcionam várias ontologias em resposta a uma determinada busca, sendo assim uma visão integrada serviria de suporte a melhoria das ontologias existentes e ao desenvolvimento de novas ontologias. Por exemplo, ao realizar um consulta por ontologias da área acadêmica, o Swoogle poderia retornar, além das ontologias disponíveis, uma visão integrada sobre as mesmas. Com isto, caso o desenvolvedor já tenha uma ontologia construída, a visão integrada daria suporte a avaliação, ou caso não tenha, daria suporte a construção de uma nova ontologia. Dessa forma poderia, talvez, alcançar um ontologia única para um determinado domínio.
- realizar estudos de caso mais abrangentes, considerando ontologias de domínio reais. Um exemplo onde isso poderia ser possível é sobre o conjunto de ontologias Biomédicas (OBO). Acredita-se que a partir de uma ontologia global gerada com base

na proposta deste trabalho, será facilitada a associação de termos destas várias ontologias no processo de anotação genômica (BELLOZE, 2007). Desta forma, as descobertas genômicas passarão a ser compartilhadas mais amplamente.

8 REFERÊNCIAS BIBLIOGRÁFICAS

AKOGRIMO. **Web site oficial** [http:// www.akogrimo.org](http://www.akogrimo.org).

ALMEIDA, M. B. **Um Modelo Baseado em Ontologias para Representação da Memória Organizacional**. Tese de doutorado, UFMG, Brasil, 2006.

ANDROUTSELLIS-THEOTOKIS, S., SPINELLIS, D. **A survey of peer-to-peer content distribution Technologies**. *Journal ACM Comput. Survey* v. 36, n. 4, ISSN 0360-0300, p. 335-371, 2004.

ANGELE, J., WEITEN, M. **Semantic Technologies in the SIMDAT Grid Project**. Internationales Begegnungs und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2005.

ARENAS, M., KANTERE, V., KEMENTSIETSIDIS, A., KIRINGA, I., MILLER, R. J., MYLOPOULOS, J. **The Hyperion Project: From Data Integration to Data Coordination**. *ACM SIGMOD Record*, 32(3):53-58, 2003.

ATH. ACM topic hierarchy **Web site oficial** - <http://www.acm.org/class/1998/>

BELLAHSÈNE, Z., ROANTREE, M., KING, N. **Services for large scale P2P Networks**. Proceedings of ERCIM Working Group on the Semantic Web, 2004. Disponível: <http://www.computing.dcu.ie/~nking/ercim04.pdf> [Capturado em 10 set. 2007].

BELLAHSÈNE, Z., LAZANITIS, C., MCBRIEN, P. M., RIZOPOULOS, N. **iXPeer: Implementing layers of abstraction in P2P Schema Mapping using AutoMed**. 2nd Workshop on Innovations in Web Infrastructure, Edinburgh, England, 2006. Disponível: <http://pubs.doc.ic.ac.uk/ixpeer-layers-of-abstraction-p2p/ixpeer-layers-of-abstraction-p2p.pdf> [Capturado em 10 jul. 2007].

BERNERS-LEE, T., HENDLER, J., LASSILA, O. **The Semantic Web**. *Scientific American*, 2001. Disponível: <http://www.sciam.com/article.cfm?id=the-semantic-web> [Capturado em: 27 nov. 2007]

- BOUQUET, P., SERAFINI, L., ZANOBINI, S. **Semantic coordination: a new approach and an application.** In Proc. of the 2nd International Semantic Web Conference (ISWO'03). Sanibel Islands, Florida, USA, October, 2003. Disponível: <http://citeseer.ist.psu.edu/bouquet03semantic.htm> [Capturado em: 10 nov. 2007].
- BOYD, M., KITTIVORAVITKUL, S., LAZANITIS, C., MCBRIEN, P. J., RIZOPOULOS, N. **AutoMed: A BAV Data Integration System for Heterogeneous Data Sources.** In Proceedings of CAiSE04, Springer Verlag, LNCS v. 3084, p. 82-97, ISSN 0302-9743, ISBN 3-540-22151-4, 2004.
- BREITMAN, K. K. **Web Semântica: A Internet do futuro.** Ed. LTC Livros Técnicos, ISBN 85-216-1466-7, Rio de Janeiro, 2005.
- BRIGHT, M. W. & HURSON, A.R., PAKZAD, S.H. **A Taxonomy and Current Issues in Multidatabase Systems.** IEEE Computer, v. 25(3), p. 50-60, 1992.
- BRIN, S. & PAGE, L. **The Anatomy of a Large-Scale Hypertextual Web Search Engine.** Computer Networks and ISDN Systems Journal, v. 30, n. 1-7, p. 107-117, 1998. Disponível: <http://citeseer.ist.psu.edu/brin98anatomy.html> [Capturado em: 10 set. 2007].
- BRITO, G. A. D. D., MOURA, A. M. C. **Integração de Objetos de Aprendizado no ROSA - P2P.** XX SBBD, Uberlândia, MG, Brasil, 2005.
- BRITO, G. A. D. D. **Integração de objetos de aprendizagem no sistema ROSA- P2P.** Tese de Mestrado, Instituto Militar de Engenharia, IME, RJ, Brasil, 2005.
- BROEKSTRA, J., EHRIG, M., HAASE, P., HARMELEN, F. V., MENKEN, M., MIKA, P., SCHNIZLER, B., SIEBES, R. **Bibster - A Semantics-Based Bibliographic Peer-to-Peer System.** Proc. of the WWW'04 Workshop on Semantics in Peer-to-Peer and Grid Computing, 2004. Disponível: <http://citeseer.ist.psu.edu/broekstra04bibster.html> [Capturado em 23 set. 2007].
- BRUIJN, J., MARTÍN-RECUERDA, F., MANOV, D., EHRIG, M. **State-of-the-art survey on Ontology Merging and Aligning V1.** EU-IST Integrated Project (IP) IST-2003-506826 SEKT, 2004. Disponível: <http://www.aifb.uni-karlsruhe.de/WBS/meh/publications/debruijn04state.pdf> [Capturado em 12 jun. 2007].

BORST, W. N. **Construction of Engineering Ontologies for knowledge sharing and reuse.** Tese de Doutorado, PhD thesis, University of Twente, Enschede, 1997.
Disponível: <http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf> [Capturado em 23 jun. 2007].

BURANARACH, M. **A Framework for the Organization and Discovery of Information Resources in a WWW Environment Using Association, Classification e Deduction.** Tese de Doutorado, University of Pittsburgh, 2004. Disponível:
http://www.sis.pitt.edu/~dist/programs/PhD/dissertations/Buranarach_M_Dissertation_13Dec2004_revised.pdf [Capturado em 21 nov. 2007].

CASTANO, S., ANTONELLIS, V., FERRARA, A., OTTATHYCAL, G. S. **Ontology-based integration of heterogeneous xml datasources.** In: Proc. of 10th Italian Symposium on Advanced Database Systems, SEBD'02, p. 27-41, 2002.

CIRNE, W., SANTOS-NETO, E. **Grades Computacionais: da Computação de Alto Desempenho a Serviços sob Demanda.** SBRC, Brasil, 2005.

COMBECHEM. **Web site oficial** - <http://www.combechem.org>

CMS. **Web site oficial** - <http://www.aktors.org/crosi/>

CORCHO, O., LÓPEZ-FERNÁNDEZ, M., GÓMEZ-PÉREZ, A. **Methodologies, tools and languages for building ontologies. Where is their meeting point?.** Data & Knowledge Engineering, Elsevier v. 46, a. 1, p. 41-64, ISSN:0169-023X, 2003.
Disponível: <http://portal.acm.org/> [Capturado em 21 set. 2007].

DE ROURE, D., BAKER, M. A., JENNINGS, N. R., SHADBOLT, N. R. **The evolution of the Grid.** In Berman, F., Fox, G. e Hey, A. J. G., Eds. Grid computing - making the global infrastructure a reality, p. 65-100, John Wiley and Sons Ltd., 2003.

DITTRICH, K., DOMENIG, R. **Towards Exploitation of the Data Universe.** In: 3^a International Conference on Business Information System, Poznań, Poland, 1999.

DITTRICH, K. & ZIEGLER, P. **Three Decades of Data Integration-All Problems Solved?** Database Technology Research Group, Department of Informatics, University of Zurich, 2003.

DMOZ. **Web site oficial** - <http://www.dmoz.org/>

- DOU, D., MCDERMOTT, D., QI, P. **Ontology Translation on the Semantic Web.** International Conference on Ontologies, Databases and Applications of Semantics, ODBASE2003, Catania, Sicily, Italy, 2005. Disponível: http://cs-www.cs.yale.edu/homes/dvm/daml/ontomerge_odbase.pdf [Capturado em 12 jul. 2007].
- DOU, D., MCDERMOTT, D., QI, P. **Ontology translation by ontology merging and automated reasoning.** In: EKAW'02 Workshop on Ontologies for Multi-Agent Systems. Siguenza, Spain, 2002. Disponível: <http://cs-www.cs.yale.edu/homes/ddj/papers/DouEtal-MAS.pdf> [Capturado em 12 set. 2007].
- DOAN, A., MADHAVAN, J., DOMINGOS, P., HALEVY, A. **Ontology Matching: A Machine Learning Approach.** In S. Staab and R. Studer, editors, Handbook on Ontologies in Information Systems. Springer-Verlag, 2003. Disponível: <http://citeseer.ist.psu.edu/doan03ontology.html> [Capturado em 13 nov 2007].
- ECLIPSE. **Web site oficial** <http://www.eclipse.org/>
- ELMASRI, T. & NAVATHE, S. **Sistema de Banco de Dados - Fundamentos e Aplicações** - 4ª Edição Editora Pearson Education, 2005.
- EUZENAT, J. **An API for ontology alignment.** In Proceedings of the 3rd International Semantic Web Conference (ISWC), pp 698-712, Hiroshima (JP), 2004.
- FALCONER, S., M., NOY, N. F., STOREY, M. A. **Towards understanding the needs of cognitive support for ontology mapping.** Stanford University. International Workshop on Ontology Matching, Geórgia, USA, 2006. Disponível: <http://www.dit.unitn.it/~p2p/OM-2006/3-Falconer-TP-OM'06.pdf> [Capturado em 22 set. 2007].
- FARQUHAR, A., FIKES, R., RICE, R. **The Ontolingua Server: a Tool for Collaborative Ontology Construction** Int. Journal of Human-Computer Studies, 46, p. 707-727, 1996. Disponível: <http://www.citeseer.ist.psu.edu/farquhar96ontolingua.html> [Capturado em 12 nov. 2007].
- FELICÍSSIMO, C. H., **Interoperabilidade Semântica na Web: Uma Estratégia para o Alinhamento Taxonômico de Ontologia.** Dissertação de Mestrado, PUC-Rio, 2004.
- FERNANDES, D. Y. S. **Reformulação de Consultas Baseada em Semântica para PDMS.** Exame de qualificação e proposta de tese, UFPE, Brasil, 2007.

FERNANDEZ, M., GÓMEZ-PÉREZ, A., JURISTO, N. **METHONTOLOGY: From Ontological Art Towards Ontological Engineering** AAAI-97 Spring Symposium on Ontological Engineering, Stanford University, March, 24-26th, 1997.

FOSTER, I. **The Anatomy of the Grid: Enabling Scalable Virtual Organizations** Proceedings of the 1st International Symposium on Cluster Computing and the Grid (CCGRID '01), 2001.

GANGEMI, A., CATENACCI, C, CIARAMITA, M., LEHMANN, J. **Modelling Ontology Evaluation and Validation** The Semantic Web: Research and Applications pp 140-154, 2006. Disponível: <http://www.springerlink.com/content/v14758128k763018> [Capturado em 2 jan. 2008].

GIUNCHIGLIA, F., SHVAIKO, P., YATSKEVICH M. **S-match: an algorithm and an implementation of semantic matching**. In Proceedings of ESWS'04, number 3053 in LNCS, p. 61.75, Heraklion, Greece, Springer-Verlag, 2004.
Disponível: [http:// http://eprints.biblio.unitn.it/archive/00000531/](http://eprints.biblio.unitn.it/archive/00000531/) [Capturado em 15 nov. 2007].

GRUBER, T. R. **Toward Principles for the Design of Ontologies Used for Knowledge Sharing**. Stanford University. International Journal of Human-Computer Studies, 1995.
Disponível: <http://citeseer.ist.psu.edu/gruber93toward.html> [Capturado em 14 set. 2007].

GRUBER, T. R. **A translation approach to portable ontology specifications**. Knowledge Acquisition. v5, i2, 199-220, 1993.

GRÜNINGER, M., FOX, M. S. **Methodology for the Design and Evaluation of Ontologies**. IJCAI-95, Workshop on Basic Ontological Issues in Knowledge Sharing, 1995.

GUARINO, N. **Understanding, Building, and Using Ontologies**. International Journal of Human and Computer Studies. v. 46, n. 2-3, p. 293-310, 1997.
Disponível: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/guarino/guarino.html> [Capturado em 12 set. 2005].

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, p. 3-15. IOS Press., 1998.
Disponível: <http://citeseer.ist.psu.edu/guarino98formal.html> [Capturado em 12 nov. 2007].

HAKIMPOUR, F., GEPPERT, A. **Resolving semantic heterogeneity in schema integration: An ontology base approach.** In Chris Welty and Barry Smith, editors, Proceedings of International conference on Formal Ontologies in Information Systems FOIS'01. ACM Press, October, 2001.
Disponível: <http://citeseer.ist.psu.edu/hakimpour01resolving.html> [Capturado em 17 nov. 2007].

HALEVY, A. Y., ZACHARY, G. I., MORK, P., TATARINOV, I. **Piazza: data management infrastructure for semantic web applications** Proceedings of the 12th international conference on WWW p. 556 - 567 ISBN:1-58113-680-3, 2003.
Disponível: <http://portal.acm.org/citation.cfm?id=775152.775231> [Capturado em 17 set. 2007].

JENA. **Web site oficial** <http://jena.sourceforge.net/>

JUDE. **Web site oficial** - <http://jude.change-vision.com/jude-web/index.html>.

KALFOGLOU, Y., HU, B. (). **CROSI Mapping System (CMS)** Results of the 2005 Ontology Alignment Contest, 2005. Disponível: <http://users.ecs.soton.ac.uk/yk1/CMS-OAEI-results.pdf> [Capturado em 10 maio de 2007].

KAON. **Web site oficial** - <http://kaon.semanticweb.org/>

BELLOZE, K. T. **Uma Extensão do Processo de Anotação Genômica para Ampliar o Uso e a Evolução Colaborativa de Ontologias no Domínio da Biologia Molecular.** Dissertação de Mestrado, Instituto Militar de Engenharia, 2007.

KOTIS, K., VALARAKOS, A., VOUIROS, G. **AUTOMS: Automating Ontology Mapping through Synthesis of Methods.** OAEI (Ontology Alignment Evaluation Initiative) contest, Ontology Matching International Workshop, Atlanta, USA, 2006.

LACHTIM, F. A., SOUZA, H. C. J., CAVALCANTI, M. C. R., MOURA, A. M. C. **Interoperabilidade de ontologias: tipos e ferramentas.** Relatório Técnico MSC 03/2008 ISSN: 1982-9035. IME - Instituto Militar de Engenharia, RJ, 2008.

LACHTIM, F. A. **Organização e Instanciação Automática de Conteúdos em Portais Semânticos.** Dissertação de Mestrado. Instituto Militar de Engenharia, RJ, 2008.

LASSILA, O., MCGUINNESS, D. **The Role of Frame-Based Representation on the Semantic Web.** Knowledge Systems Laboratory Report KSL-01-02, Stanford University, 2001. Disponível: <http://www.ep.liu.se/ea/cis/2001/005/> [Capturado em 14 set. 2007].

MAEDCHE, A., STUMME, G. **FCA-MERGE: Bottom-Up Merging of Ontologies.** Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI p. 225-234, 2001. Disponível: <http://citeseer.ist.psu.edu/392023.html> [Capturado em 12 set. 2007].

MAEDCHE, A., NUNO SILVA, A. P., MOTIK, B., VOLZ, R. **MAFRA - A Mapping FRamework for Distributed Ontologies.** Univ. Karlsruhe. Proc. of the 13th European Conf. on Knowledge Engineering – Springer, 2002. Disponível: http://www.fzi.de/KCMS/kcms_file.php?action=link&id=33 [Capturado em 12 set. 2007].

MILOJICIC, D. S., KALOGERAKI, V., LUKOSE, R., NAGARAJA, K., PRUYNE, J., RICHARD, B. ROLLINS, S., XU, Z. **Peer-to-peer computing.** Technical Report HPL-2002-57, HP Lab, 2002. Disponível: <http://citeseer.ist.psu.edu/milojicic02peertopeer.html> [Capturado em 12 set. 2007].

MIRANDA, M. G. **Recuperação de Informações em Sistemas Ponto-a-Ponto Emergentes.** Qualificação de doutorado submetida a COPPE-URFJ, 2006.

MITRA, P., NOY, N. F., JAISWAL, A. R. **Omen: A probabilistic ontology mapping tool.** In Workshop on Meaning coordination and negotiation at the Third International Conference on the Semantic Web (ISWC), Hiroshima, Japan, 2004. Disponível: <http://smi.stanford.edu/smi-web/reports/SMI-2004-1126.pdf> [Capturado em 12 nov. 2007].

NAPSTER. **WEB site oficial** - <http://www.napster.com>

NG, W. S., OOI, B. C., TAN, K., ZHOU, A. **PeerDB: A P2P-based System for Distributed Data Sharing.** In Proc. of the International Conference on Data Engineering 2003 (ICDE'03). 633-644, 2003. Disponível <http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/8910/28179/01260827.pdf?arnumber=1260827> [Capturado em 9 set. 2007].

NOY, N. F., MCGUINNESS, D. L. **Ontology development 101: a guide to create your first ontology.** Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 e Stanford Medical Informatics Technical Report SMI-2001-0880, 2001. Disponível: <http://www-ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.doc> [Capturado em 9 set. 2007].

NOY, N., MUSEN, M. **PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment.** In Proceedings of AAAI-2000, Austin, Texas, MIT Press/AAAI Press, 2000. Disponível: <http://citeseer.ist.psu.edu/noy00prompt.html> [Capturado em 10 dez. 2007].

NOY, N. F., MUSEN, M. A. **The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping.** Stanford University. International Journal of Human-Computer Studies v. 59, a. 6, p. 983-1024, 2003.

NUNO SILVA, A. P. **Multi-Dimensional Service-Oriented Ontology Mapping.** Tese de Doutorado, Universidade de Trás-os-Montes e Alto Douro, Portugal, 2004.

OAEI. *Ontology Alignment Evaluation Initiative - Web site oficial*
<http://oei.ontologymatching.org/>

OBO. **Web site oficial** <http://www.obofoundry.org/>

OMELAYENKO, B. **RDFT: A mapping meta-ontology for business integration.** In Proceedings of the Workshop on Knowledge Transformation for the Semantic Web (KTSW2002) at the 15-th European Conference on Artificial Intelligence, pages 76.83, Lyon, France, 2002.
Disponível: <http://www.citeseer.ist.psu.edu/omelayenko02rdft.html> [Capturado em 12 nov. 2007].

OMELAYENKO, B. **Integrating vocabularies: Discovering and representing vocabulary maps.** In Proceedings of the First International Semantic Web Conference (ISWC2002), Sardinia, Italy, 2002.
Disponível: <http://www.citeseer.ist.psu.edu/699976.html> [Capturado em 12 nov. 2007].

OMV. **Web site oficial** - <http://omv.ontoware.org/>

ONTOGRID. **Web site oficial** - <http://www.ontogrid.net>

OOI, B., SHU, Y., TAN, K. **Relational Data Sharing in Peer-based Data Management Systems.** ACM SIGMOD Record, 32(3), 2003.

PALMA, R., HAASE, P., GÓMEZ-PÉREZ, A. **Oyster - Sharing and Re-using Ontologies in a Peer-to-Peer Community.** Proceedings of the 15th international conference on WWW, 2005. Disponível: <http://portal.acm.org> [Capturado em 12 set. 2007].

PINTO, H., S., GOMEZ-PEREZ, A., MARTINS, J., P. **Some issues on ontology integration.** In Proceedings of the IJCAI'99 Workshop on Ontology and Problem-Solving Methods: Lesson learned and Future Trends, ed., V.R. Benjamins, volume 18, pp. 7.1--7.11, Amsterdam. CEUR Publications, 1999.
Disponível: <http://www.citeseer.ist.psu.edu/pinto99some.html> [Capturado em 12 nov. 2007].

PIRES, C. E. S. **Um sistema de Gerenciamento Dados com Conectividade Baseada em Semântica.** Exame de qualificacao e proposta de tese, UFPE, Brasil, 2007.

PROTEGE. **Web site oficial** - <http://protege.stanford.edu>

RDF. **Web site oficial** - <http://www.w3.org/RDF/>

RDFS. **Web site oficial** - <http://www.w3.org/TR/rdf-schema/>

SHETH, P. **Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics.** Interoperating Geographic Information Systems. Edited by Good-child M. & Egenhofer, M. & Fegeas, R. & Kottman, C. Kluwer Academic Publishers, pp. 5-29, 1999.
Disponível: <http://lstdis.cs.uga.edu/library/download/S98-changing.pdf> [Capturado em 12 jan. 2008].

SHETH, P., LARSON, J.A. **Federated Databases Systems for Managing Distributed Heterogeneous and Autonomous Databases.** ACM Computing Surveys (CSUR), v.22 n.3, p.183-236, Sept. 1990.

SPARQL. **Web site oficial** <http://www.w3.org/TR/rdf-sparql-query>

SURE, Y., BLOEHDORN, S., HAASE, P., HARTMANN, J., OBERLE, D. **The SWRC Ontology- Semantic Web for Research Communities.** Lv. Progress in Artificial Intelligence v. 3808/2005 ISBN 978-3-540-30737-2, c. 4, p. 218-231, 2005.

SURE, Y., ERDMANN, M., ANGELE, J., STAAB, S., STUDER, R., WENKE, D. **OntoEdit: Collaborative ontology development for the semantic web.** Proceedings of the first International Semantic Web Conference (ISWC), 2002.

SWOOGLE. **Web site oficial** <http://swoogle.umbc.edu/>

TALIA, D., TRIUNFIO, P. **Toward a Synergy Between P2P and Grids.** IEEE Internet Computing 7(4): 94-96, 2003.

USCHOLD, M., KING, M. **Towards a Methodology for Building Ontologies.** Workshop on Basic Ontological Issues in Knowledge Sharing in the IJCAI-95-International Joint Conference on AI, 1995.
Disponível em: < <http://citeseer.ist.psu.edu/uschold95toward.html>>. Acesso em: set. 2007.

VALDURIEZ, P., PACITTI, E. **Data Management in Large-Scale P2P Systems.** lv.VECPAR pg.104-118, ISBN 3-540-25424-2, 2004.

XIAO, H., CRUZ, I. F. **Ontology-based query rewriting in peer-to-peer networks.** In Proceedings of the 2nd International Conference on Knowledge Engineering and Decision Support, p. 11–18, 2006.

XIAOMENG, S. **Semantic Enrichment for Ontology Mapping.** Tese de doutorado, Norwegian University of Science and Technology, Faculty of Information Technology, Mathematics and Electrical Engineering, Department of Computer and Information Science, 2004.
Disponível: <http://www.idi.ntnu.no/grupper/su/publ/phd/xiaomeng-su-thesis.pdf>
[Capturado em 18 set. 2007].

WORDNET. **Web site oficial** - <http://wordnet.princeton.edu/>

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)