

Interoperabilidade de Conteúdos em Portais Semânticos

Ana Maria de Carvalho Moura

1. Introdução

Nos últimos anos as organizações corporativas e setores públicos de vários países vêm sofrendo fortes pressões quanto à necessidade de revitalizarem suas administrações, tornando-as mais pró-ativas, integradas, eficientes, transparentes e, especialmente, mais orientadas a serviços. Apesar do grande desafio que essa mudança de paradigma representa, estes setores vêm introduzindo gradativamente inovações em sua estrutura organizacional de forma a mobilizar, implantar e utilizar o capital humano e de informação, recursos tecnológicos e financeiros para a entrega de serviços aos cidadãos em todo o mundo (United Nations 2008). Nesse contexto, as iniciativas de governo eletrônico (e-Gov) vêm assumindo um papel essencial na sociedade como um todo: permite a divulgação de informações e a interação direta do cidadão com políticas de governo, além de proporcionar possibilidades de comércio e negócios na Web.

E-Gov refere-se, portanto, ao uso da Web e outras tecnologias de informação pelo corpo governamental para interação com seus cidadãos. A prática de gestão do conhecimento aliada à adequada utilização das Tecnologias de Informação e Comunicação (TIC) desempenha um papel crucial em iniciativas privadas e públicas, tornando-as mais eficientes e competitivas. Um ótimo exemplo de iniciativa pública no Brasil é a arquitetura do e-Ping¹ (Padrões de Interoperabilidade de Governo Eletrônico), cujo objetivo é definir um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização de TIC no governo federal, estabelecendo as condições de interação com os demais poderes e esferas de governo e com a sociedade em geral. Nesta arquitetura, fica evidenciada a forte preocupação em oferecer portais intuitivos e de fácil acesso ao cidadão em todas as esferas do governo.

Apesar do grande avanço das iniciativas públicas e privadas no uso de várias modalidades das TICs na construção de seus artefatos, raras são as iniciativas que consideram a Web Semântica como uma forte tecnologia de integração e reuso de recursos (Berners-Lee et al. 2001). A Web semântica é hoje um dos objetivos de longo prazo do W3C (W3C), com o aporte de resultados de pesquisas importantes, que contribuem com inovações e resultados tecnológicos significativos em vários centros de pesquisa no mundo. Ela deverá se desenvolver num ambiente de acesso inteligente à informação heterogênea e distribuída, permitindo um melhor uso dos recursos que estão sendo colocados à disposição da Web, tais como os agentes de softwares, tão importantes para o desenvolvimento de aplicações. Estes agentes irão realizar a mediação (*brokering*) entre as necessidades de cada usuário e as fontes de informação disponíveis, permitindo recuperações mais acuradas e eficientes. Essa mediação é feita geralmente por meio de interfaces Web, também denominadas portais, que reúnem uma variedade de serviços e conteúdos voltados para um grupo de usuários públicos ou corporativos.

¹<http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padroes-deinteroperabilidade>.

Portal é um site Web que coleta informações para um grupo de usuários que possuem interesses em comum (Toledo 2002), sendo este um ponto de entrada para a localização de conteúdos e realização de serviços de interesse de uma comunidade. No contexto de e-gov, serviços podem ser categorizados como: Governo para Cidadãos (G2C), Governo para Negócios (G2B), Governo para Empregados (G2E) e Governo para Governo (G2G) (Saldhana 2007). Porém, independentemente do objetivo ao qual se destinam, todos devem prover requisitos de total segurança de forma a conquistar a confiança plena de seus usuários, bem como prover mecanismos de comunicação que permitam trocar e intercambiar conteúdos entre outros portais.

2. Motivação e objetivos

Nesse contexto, foi desenvolvido no IME-RJ, um projeto de pesquisa que teve como objetivo demonstrar que o uso de tecnologias da Web Semântica contribui sobremaneira para uma melhoria na organização, integração e gestão das informações em portais. Em (Lachtim et al. 2008a) (Lachtim 2008) foi desenvolvida uma arquitetura que serviu como infra-estrutura básica para a construção de Portais Semânticos, capaz de integrar e instanciar informações a partir do uso intensivo de ontologias (Maedche 2001). Essa arquitetura permite que informações distribuídas na Web sejam recuperadas com base em uma ontologia de domínio e, através de mapeamentos definidos entre essas informações e a ontologia de domínio, novos conteúdos sejam dinamicamente categorizados e adicionados ao portal.

Uma característica interessante contemplada por essa arquitetura consiste na adoção do conceito de facetar como estratégia de organização de conteúdos desses portais. Além de reduzir o tempo de busca às informações relevantes na Web, essa estratégia ajuda usuários inexperientes, já que sites e documentos são agrupados em categorias pré-definidas, de acordo com seus conteúdos. A idéia de classificação multifacetada, proposta em (Ranganathan e Gopinath 1967), tem sido empregada pela comunidade de pesquisa em recuperação de informação e encontrada sob várias formas na Web. Essa prática deu origem à terminologia busca facetada semântica, onde facetar são baseadas em estruturas ontológicas dotadas de capacidade de raciocínio (Mäkelä et al. 2005), uma idéia que já vem sendo explorada em portais semânticos (Suominen et al. 2007) (Sidoroff et al. 2005).

De forma a evidenciar a aplicabilidade dessa arquitetura, foi desenvolvido o PORTAL SEMântico EDUCacional (POSEDU²) (Ferreira e Gama, 2008), (Lachtim et al. 2008b) contemplando dois módulos importantes da arquitetura proposta em (Lachtim 2008): o módulo de organização e instanciação do portal e o módulo de mineração, também denominado de KDD (Knowledge Data Discovery) (Frawley 1991). O primeiro módulo é responsável pela organização e alimentação automática do portal. O segundo módulo implementa uma alternativa complementar para a obtenção de documentos de um domínio específico, utilizando-se para isto de um mecanismo de busca na Web aberta (Yahoo). Os documentos retornados são então agrupados e classificados com base nessa ontologia e inseridos no portal como links recomendados nas facetar correspondentes.

No entanto, ao longo desse trabalho, observou-se um número muito restrito de ontologias na Web, mesmo em domínios de grande relevância, a exemplo de educação. A grande maioria dos portais educacionais das universidades e centros de pesquisa tanto no Brasil quanto no exterior não fazem uso de ferramentas de anotação para publicação

² www.comp.ime.eb.br/~posedu

de suas páginas e documentos. Anotações são descrições atribuídas (metadados) aos recursos publicados na Web. Além de simples metadados, anotações permitem agregar conhecimento semântico adicional aos recursos associando, por exemplo, propriedades e restrições aos quais, denominamos ontologias (Gruber 1993), contribuindo para a recuperação dos mesmos com base em suas características semânticas. Na literatura existem algumas de ferramentas de anotação desenvolvidas para este fim, a exemplo de Amaya³, Gate⁴, Annotea⁵, Saha (Valkeapää et al. 2007), Magpie (Domingue e Dzbor 2004), etc., porém ainda pouco difundidas entre os usuários Web.

Em (Lachtim 2008) verificou-se que algumas instituições de ensino, a exemplo das universidades de Lehigh⁶ (Pensilvânia, Estados Unidos) e Munique⁷ (Alemanha), apresentam suas estruturas organizacionais através de taxonomias representadas em OWL, porém com poucas ou quase nenhuma instância. Portanto, observa-se que o uso de ferramentas de anotação ainda é bastante insipiente.

Dando continuidade a essa pesquisa, o objetivo principal desse projeto é ampliar a utilização de portais semânticos no âmbito de governo eletrônico e permitir a integração de conteúdos entre outros portais de forma a complementar o escopo de informações buscadas pelo usuário do portal.

Para tanto é necessário inicialmente ampliar o acesso a ontologias em diversos domínios do conhecimento, a exemplo de Saúde, Medicina, Educação, Esportes, etc. de modo que portais semânticos possam interagir e compartilhar informações a partir da Web. Nesta primeira etapa do trabalho (1º ano) serão estudados os requisitos e tecnologias necessárias para a extração de conteúdos da Web, tendo como base taxonomias já publicadas em sites populares e governamentais, de modo a criar ontologias em diversos domínios. Numa segunda etapa do trabalho (2º ano), essas ontologias serão armazenadas num repositório de metadados para integração de conteúdos em portais semânticos.

3. Ontologias e Regras Semânticas

Ontologia é definida como um conjunto de conceitos e termos que podem ser usados para descrever alguma área do conhecimento ou construir uma representação para o conhecimento.

Uma ontologia pode ser descrita conforme a especificação formal definida em (Ehrig e Sure 2004) como $O := (C, HC, RC, HR, RI, I, A)$, onde uma ontologia O é uma tupla que consiste de: Conceitos C , que podem ser arranjados em uma hierarquia HC . As relações RC (ou propriedades) entre conceitos também podem ser arranjadas em uma hierarquia HR . Instância I de um conceito específico é interconectada por instâncias de propriedades RI . Adicionalmente, uma ontologia pode definir axiomas A , usados para inferir conhecimento. Dessa forma, uma ontologia é representada por um conjunto de termos hierarquicamente estruturados para descrever um domínio que pode ser usado como um esqueleto fundamental para uma base de conhecimento. Conceitos representam um conjunto ou classes de entidades; relações representam o tipo de interação entre os conceitos de um domínio; axiomas contêm regras, restrições de

3 www.w3.org/Amaya/

4 <http://gate.ac.uk/gate/doc/releases.html>

5 www.w3.org/2001/Annotea/

6 <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl>

7 <http://www.radiig.in.tum.de/ontology/organisation>

valores para classes ou instâncias; e instâncias são utilizadas para representar elementos específicos, ou seja, os próprios dados. Assim, ontologias podem ser usadas para:

- Explicitar significado aos objetos que nos rodeiam;
- Integrar recursos na Web permitindo a interoperabilidade semântica;
- Fornecer um entendimento comum dos conceitos para a comunicação entre agentes;
- Integrar esquemas de diferentes bases de dados.

Nos portais semânticos o conhecimento é representado por meio de ontologias, permitindo que este possa ser compartilhado e utilizado por outras aplicações e agentes de software, auxiliando as pessoas nos processos de buscar, extrair, interpretar, instanciar e processar a informação (LAUSEN et al., 2005), tarefas deixadas à mercê do usuário nos portais tradicionais. Outra grande vantagem dos portais semânticos é a possibilidade de declarar e processar axiomas, o que possibilita a inferência de novos conteúdos e sua agregação à base de conhecimento dos portais.

Considere, por exemplo, uma ontologia de origem acadêmica, contendo termos "universidade", "publicação", "pessoa" e "autor", e um conjunto de axiomas ou regras com declarações do tipo: "todos os artigos são publicações" ou "os autores das publicações são pessoas". A partir destas declarações, é possível inferir novos fatos, como por exemplo, as "publicações realizadas por um determinado autor". Este tipo de informação ajuda sobremaneira no processo de recuperação de resultados de buscas dos portais, até então inviáveis de serem obtidos nos sistemas de recuperação convencionais (Lausen et al., 2005). Linguagens de regras, a exemplo do SWRL (Semantic Web Rule Language) (Horrocks et al. 2004) também têm sido utilizadas como um importante mecanismo de suporte à inferência de dados, complementando a lacuna existente no OWL no tocante construção de axiomas complexos. Estas linguagens permitem compor regras do tipo $X: Y \rightarrow Y$, onde X e Y representam respectivamente a cabeça e o corpo da regra. O grande diferencial dessas linguagens em relação ao OWL é que o corpo da regra permite a composição entre relacionamentos, não necessariamente dentro da mesma hierarquia, o que não é possível em OWL-DL.

4. Portais Semânticos e Interoperabilidade

Os portais vêm constantemente evoluindo de modo a automatizar muitos dos processos envolvidos em sua utilização e manutenção, e uma das vertentes que vêm contribuindo para essa evolução é a Web Semântica. Portais amparados por esta modalidade são denominados Portais Semânticos, que se caracterizam por armazenarem e estruturarem conteúdos informacionais segundo ontologias específicas de domínio (Lausen et al., 2005), (Mäkelä et al 2004), (Corcho et al. 2003). Sendo uma evolução dos portais tradicionais, os portais semânticos estabelecem um novo padrão para a busca, processamento, organização, compartilhamento e inferência de informações. Seu desenvolvimento está centrado no uso intensivo de ontologias, objetivando permitir o compartilhamento e o entendimento comum de um domínio entre pessoas e sistemas de aplicação, viabilizando também estrutura e significado semântico aos documentos acessados pelo portal, armazenando-os em seus repositórios ou na Web (Pinheiro e Moura 2004).

Portanto, além de se basear em tecnologias Web já utilizadas em portais convencionais, os portais semânticos fazem uso de linguagens de representação de

taxonomias e ontologias como RDF⁸ e OWL⁹. Tal incremento contribui para a utilidade prática dos portais, pelo fato de que uma ontologia se beneficia de uma terminologia adequada para descrever conteúdos e axiomas formais usados na criação de novas relações, e restringe suas interpretações segundo um sentido pretendido (Noy e Hafner 1997).

Nos últimos anos inúmeros esforços vêm se concentrando no desenvolvimento de arquiteturas e gerenciadores de portais semânticos, a exemplo do *SWED* (Brickley et al. 2002), *OntoWeb* (Jarrar et al. 2002), *OntoViews* (Mäkelä et al. 2004), *PASS* (Pinheiro e Moura, 2004), *SiGePoS* (Lachtim et al. 2008a) e *ODESeW* (Corcho et al., 2003).

No contexto deste projeto o foco maior é dado ao serviço G2C, onde conteúdos de um portal semântico são voltados para suprir a curiosidade de usuários em um determinado domínio de conhecimento. Assim, é imprescindível que estes possam recuperar informações, quer do próprio portal, quer da integração com outros portais de domínios similares ou complementares.

Interoperabilidade é um requisito fundamental para o desenvolvimento de serviços e-Gov eficientes e efetivos (Commission Staff 2003). Em essência, a interoperabilidade nos portais envolve 3 aspectos distintos:

- Interoperabilidade técnica: trata das questões técnicas envolvendo a ligação entre sistemas de computadores, definição de interfaces abertas, formatos de dados e protocolos e telecomunicações;
- Interoperabilidade semântica: trata dos mecanismos capazes de assegurar que as informações trocadas entre aplicações sejam perfeitamente entendidas;
- Interoperabilidade organizacional: trata da modelagem dos processos de negócio, alinhando informações de arquiteturas com objetivos organizacionais, auxiliando esses processos a cooperarem entre si.

Na prática, a integração de serviços é um requisito básico dos portais administrativos, que buscam juntar e transformar processos de interesse de um cidadão.

Nesse projeto o alvo de maior interesse é voltado para a interoperabilidade semântica, por meio da qual conceitos entre ontologias complementares ou similares possam ser comparadas e, em caso de semelhança, suas instâncias possam ser integradas ou mapeadas entre si.

4.1 Interoperabilidade Semântica

Um dos objetivos propostos no escopo deste projeto é integrar conteúdos a partir de outros portais, o que significa capturar e reutilizar informações de outras páginas segundo um determinado contexto.

Reutilizar informações de diferentes ontologias requer mecanismos específicos e eficientes, capazes de lidar com diferentes níveis de interoperabilidade (Felicíssimo 2004). Quatro abordagens distintas para lidar com interoperabilidade entre ontologias são propostas na literatura: mapeamento, alinhamento, fusão ou merge e integração (Noy et al. 2006).

Mapeamento de ontologias determina quando duas ontologias estão semanticamente relacionadas no nível conceitual, e como instâncias de uma ontologia

8 www.w3.org/RDF/

9 www.w3.org/2004/OWL/

fonte são transformadas em instâncias da ontologia alvo, de acordo com suas associações semânticas, provendo também mecanismos para transferir instâncias entre elas (Maedche et al 2002). Mais voltado para lidar com ontologias complementares, o alinhamento resulta em um acordo mútuo, onde duas ontologias fonte tornam-se consistentes e coerentes, gerando como resultado final as ontologias originais separadas, mas com links adicionais entre seus termos equivalentes.

Fusão ou Merge é o processo de unificação de ontologias de domínios similares, onde versões das ontologias originais são combinadas em uma única, com todos os seus termos e sem referência às suas origens.

Finalmente, na integração três perspectivas podem ser consideradas (Pinto et al 1999): (1) a abordagem de merge, tal como descrita anteriormente; (2) quando uma aplicação usa conceitos de uma ou mais ontologias; e (3) quando uma ontologia de domínio é criada a partir do reuso de conceitos de uma ou mais ontologias de domínios diferentes. Reuso aqui é caracterizado como o resultado da generalização, especialização ou adaptação de conceitos de outras ontologias existentes. Essa última perspectiva é a que consideramos no contexto desse trabalho.

Em resumo, essas abordagens podem ser consideradas como um processo de reuso de ontologias, e apresentam diferentes níveis de comprometimento no que diz respeito à comparação entre ontologias: alinhamento < mapeamento < integração < merge. O alinhamento estabelece apenas correspondências entre entidades de ontologias, enquanto no mapeamento existe uma infra-estrutura para transferir instâncias de uma ontologia fonte para uma ontologia alvo (Maedche et al 2002), (Noy e Musen 2002). Já no processo de integração uma nova ontologia é criada a partir do reuso de outras, muito embora mantenha as referências para as ontologias fonte. Finalmente, a abordagem *merge* gera uma nova ontologia sem preservar nenhum link com as ontologias originais.

Entretanto, todas essas abordagens são precedidas por uma técnica denominada na literatura como casamento (ou *matching*) (Euzenat e Shvaiko 2007), (Hu e Qu 2008), (Jean-Mary e Kabuka 2007), (Kalfoglou et al. 2005), embora outros trabalhos também considerem essa técnica como alinhamento (Euzenat e Valtchev 2004), (Euzenat et al. 2008) (Jean-Mary e Kabuka 2007) (Kiu e Lee 2007), ou mesmo como mapeamento (Nagy et al. 2007). Independentemente do nome, essa técnica consiste em comparar termos de uma ontologia alvo com os termos de uma ontologia fonte, de modo que medidas de similaridade léxica, estrutural e semântica possam ser definidas entre esses termos. Assim, quanto mais refinados forem os resultados de similaridade retornados pelo processo de casamento, melhores serão os resultados fornecidos pelo mecanismo de interoperabilidade.

Nos últimos anos a literatura aponta várias ferramentas desenvolvidas para automatizar o processo de casamento de ontologias (Euzenat e Shvaiko 2007) (Euzenat et al. 2008), embora utilizando diferentes estratégias e algoritmos: ASMOV (Jean-Mary e Kabuka 2007), Falcon-AO (Hu e Qu 2008), Ontodna (Kiu e Lee 2007), DSSim (Nagy et al. 2007), CMS (CROSI Mapping System) (Kalfoglou et al. 2005) são alguns desses exemplos.

Em (Latchim 2008) partimos do CMS para realizar o casamento de ontologias. Porém, devido a algumas de suas limitações desenvolvemos o *e-CMS* (Latchim et al. 2009) que trouxe um benefício significativo ao processo de casamento.

Nesse projeto pretendemos realizar a extração de conteúdos para a criação de

ontologias de domínio utilizando técnicas de mineração em texto de páginas Web associadas a técnicas de descoberta de conhecimento na Web profunda. A Web profunda caracteriza-se como aquela que armazena informações não detectadas por robôs, a exemplo de formulários HTML ou banco de dados (B. He, et al. 2007). Diversos trabalhos têm sido realizados com o objetivo de definir mecanismos e algoritmos visando extrair informações (estruturadas ou não) relevantes da Web profunda e de *sites* sociais (Berti- L. Equille et al, 2009), (J. Madhavan et al, 2009), (A. Doan et al, 2009), (S. Amer-Yahia, 2009 et al, 2009), (M. Cafarella, 2009), (Barbosa e Freire, 2007).

5. Amadurecimento do Projeto no Grupo de Pesquisa

A professora proponente tem se dedicado nos últimos 10 anos ao desenvolvimento de trabalhos de pesquisa e à preparação de material didático na linha de Web Semântica. Foi responsável pela orientação de mais de 15 dissertações, com resultados publicados em periódicos, conferências nacionais e internacionais, além de ter participado e coordenado projetos de pesquisa na área (Moura 2009).

Tendo se aposentado há cinco anos do IME-RJ, permanece como professora colaboradora junto a essa instituição, onde atualmente desfruta de uma bolsa FAPERJ como pesquisadora visitante (proc. E-26/151.304/2007). Trabalhou também durante três anos junto ao Proderj (Centro de Tecnologia da Informação e Comunicação do Estado do Rio de Janeiro), onde realizou atividades de consultoria e planejamento em governo eletrônico.

Os resultados obtidos ao longo de sua vida acadêmica e de pesquisa demonstram, portanto, a experiência e o amadurecimento requeridos para a realização de um projeto dessa natureza.

6. Plano de Trabalho e Cronograma

A principal meta nesse trabalho é aprofundar no estudo das tecnologias de interoperabilidade de ontologias visando a integração de conteúdos em portais semânticos. Esse trabalho, elaborado para um período de 2 anos é constituído das seguintes etapas:

- I. Extração de conteúdos da Web (rasa e profunda) em um dado domínio, partindo de taxonomias bem definidas em sites populares (Yahoo, Google, ...) e governamentais (Paraná, Rio de Janeiro, ...). Nesta etapa serão empregadas técnicas de mineração de texto associadas a algoritmos de clusterização e classificação para seleção de conteúdos relevantes.
- II. Construção de ontologias de domínio a partir dos conceitos das taxonomias e conteúdos extraídos.
- III. Criação de um repositório de metadados para acesso a essas ontologias.
- IV. Integração do repositório a um portal semântico, com vistas à adição de novos conteúdos. O projeto contempla o uso do portal semântico gerado através do sistema SiGePoS (Latchim 2008).

No entanto, no contexto do projeto em questão definido para um período de um ano, pretendemos trabalhar no item I e iniciar a etapa II. A primeira etapa apresenta um grau de complexidade elevado devido às inúmeras técnicas disponíveis na literatura para extração de conteúdos na Web profunda, que precisam ser estudadas e avaliadas para o sucesso da etapa II.

Um plano mais detalhado para o desenvolvimento das etapas I e II é listado a seguir.

- i) Proceder ao levantamento detalhado de requisitos nas áreas alvo do projeto.
- ii) Definição do escopo de domínio comum que irá contribuir com os requisitos de especificação e para os quais concentrar-se-ão nossos experimentos e aplicações.
- iii) Estudo e seleção de taxonomias
- iv) Estudo das principais técnicas disponíveis na literatura para a descoberta de dados na Web. Essa etapa deverá ser acompanhada de testes que comprovem a eficácia dos mecanismos e como poderão ser aplicados ao trabalho proposto.
- v) Extração de conteúdos segundo a taxonomia escolhida
- vi) Definição e especificação do módulo de criação de ontologias.

Atividades	1	2	3	4	5	6	7	8	9	10	11	12
Levantamento detalhado de requisitos nas áreas alvo do projeto	█	█										
Definição do escopo de domínio		█	█									
Estudo e seleção de taxonomias			█	█								
Estudo das principais técnicas para a descoberta de dados na Web		█	█	█	█	█	█					
Extração de conteúdos								█	█	█		
Definição e especificação do módulo de criação de ontologias									█	█	█	█

7. Resultados Esperados

Os resultados de pesquisa serão produzidos no contexto de uma dissertação de mestrado e servirão como referência para conteúdos de disciplinas envolvendo a área de Banco de Dados ministradas no LNCC. Adicionalmente, ao longo do trabalho, serão elaborados artigos de referência publicados em periódicos e conferências internacionais. Os protótipos e técnicas desenvolvidos poderão ser utilizados pela comunidade acadêmica e instituições governamentais em geral, no sentido de facilitar o desenvolvimento de aplicações de governo eletrônico.

Referências

- S. Amer-Yahia, L. Lakshmanan, C. Yu. SocialScope:Enabling Information Discovery on Social Content Sites (2009). *4th Biental Conference on innovative data systems (CIDR)*, Asilomar, CA, EU, jan.
- L. Barbosa, J. Freire (2007). An Adaptative Crawler for Locating Hidden-Web Entry Points.
- Berners-Lee, T., Hendler, J., E Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5):34-43, maio.
- L. Berti-Equille, A. Sarma, X. Dong, A. Marian, D. Srivastava (2009). Sailing the Information Ocean with Awareness of Currents: Discovery and Application of Source Dependence. *4th Biental Conference on innovative data systems (CIDR)*, Asilomar, CA, EU, jan.
- D. Brickley, S. Buswell, B. M. Matthews, L. Miller, D. Reynolds, M.D Wilson (2002). Semantic Web Advanced Development for Europe (SWAD-Europe). In *Proceedings of*

the First International Semantic Web Conference on the Semantic Web, Sardinia, pages 409-413.

M. Cafarella (2009). Extracting and Querying a Comprehensive web Database. *4th Bienal Conference on innovative data systems (CIDR)*, Asilomar, CA, EU, jan.

Commission of the European Communities, Linking up Europe: the Importance of Interoperability for e-Government Services (2003). *Commission Staff Working Paper SEC(2003) 801*, July.

O. Corcho, A. Gómez-Pérez, A. López-Cima, V. López-Garcia, MC. Suárez-Figueroa (2003). ODESeW. Automatic generation of knowledge portals for intranets and extranets. In *International Semantic Web*, Florida, pages 802-817.

A. Doan, J. Naughton, A. Baid, X. Chai, et al. (2009). The Case for a Structured Approach to Managing Unstructured Data. *4th Bienal Conference on innovative data systems (CIDR)*, Asilomar, CA, EU, jan.

J. Domingue, M. Dzbor e E.Motta (2004). Magpie: Supporting Browsing and Navigation on the Semantic Web. *Proc. of the 9th Int. Conf. On Intelligent User Interfaces*, Portugal.

M. Ehrig, e Y. Sure (2004). Ontology Mapping - an Integrated Approach. *Proc. Of the 1th European Semantic Web Symposium*, volume 3053 of *Lecture Notes in Computer Science*, pags. 76-91. Springer Verlag, maio.

J. Euzenat e P. Shvaiko (2007). *Ontology Matching*. Springer-Verlag, Berlin Heidelberg.

J. Euzenat e P. Valtchev (2004). Similarity-Based Ontology Alignment in OWL-Lite. *Proceedings. of the 16th European Conference on Artificial Intelligence (ECAI'04)*, Valence (SP), 22-27 ago.

J. Euzenat, A. Isaac, C. Meilicke, P. Shvaiko, H. Stuckenschmidt³, O. Šváb⁵, V. Svátek⁵, W. R. Hage², and M. Yatskevich (2007). First Results of the Ontology Alignment Evaluation Initiative 2007. Available at <http://oaei.ontologymatching.org/2007/results/oaei2007.pdf>, last access Dec. 2008.

C. H. Felicíssimo (2004). Interoperabilidade Semântica na Web: uma Estratégia para o Alinhamento Taxonômico de Ontologias. *Dissertação de Mestrado*, PUC, Rio de Janeiro, Brasil, ago.

G. Ferreira e R. Gama (2008). Organização Automática de Páginas Web para Exibição em Portais Semânticos. Projeto de fim de curso de graduação em Eng. de Computação, IME, agosto.

W. J. Frawley, G. Piatetsky-Shapiro, C. J. Matheus, (1991). Knowledge Discovery in Databases: an Overview. *Knowledge Discovery in Databases*, AAAI/MIT Press.

T. R. Gruber (1993). Towards principles for the design of ontologies used for knowledge sharing. Em GUARINO, N. e POLI, R., editores, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, the Netherlands, Kluwer Academic Publishers.

B. He, M. Patel, Z .Zhang, K. Chang (2007). Accessing the Deep Web: a Survey. *CACM*, 50(5): 95-101.

I. Horrocks, P. F. Patel-Schneider, H. Boley, S.Tabet, B. Grosf, M. Dean (2004). SWRL: A Semantic Web Rule Language Combining OWL and RuleML, W3C Member

Submission, <http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/>, último acesso fev 2009.

W. Hu, Y. Qu (2008). Falcon-AO: A Practical Ontology Matching System. Web Semantics Sci Serv Agents World Wide Web.

M. Jarrar, B. Majer, R. Meersman, P. Spyns, R. Studer, Y. Sure, R. Volz (2004). Web Portal, Complete Ontology and Portal. *Technical Report*, Ontoweb Project (Ist-2000-29243), Institute AIFB, University of Karlsruhe, May.

Y. Jean-Mary, M. Kabuka (2007). Asmov Results for OAEI 2007. *Proceedings of the 2nd Ontology Matching Workshop*, Busan, Korea.

Y. Kalfoglou, B. Hu, D. Reynolds, and N. Shadbolt (2005). Capturing, Representing and Operationalising Semantic Integration. *Technical Report*, University of Southampton, School of Electronics and Computer Science Advanced Knowledge Technologies group, out.

C-C. Kiu, C.-S. Lee (2007). Ontodna: Ontology Alignment Results for OAEI 2007, in: *Proceedings of the 2nd Ontology Matching Workshop*, Busan, Korea.

F. Lachtim (2008). Organização e Instanciação de Conteúdos em Portais Semânticos. *Tese de Mestrado*, IME, maio.

F. Lachtim, G. Ferreira, R. Gama, A. M. Moura, M. C. Cavalcanti (2008a). An Architecture for Dynamic Organization and Publication in Semantic Portals. *10th International Conference on Information Integration and Web-Based Applications and Services (IIWAS)- Emerging Research Projects, Applications and Services Symposium (ERPAS)*, Linz, Austria, nov.

F. Lachtim, G. Ferreira, R. Gama, A. M. Moura, M. C. Cavalcanti (2008b). POSEDU: um Portal Semântico Educacional, *2nd Brazilian Workshop Semantic Web and Education (SBIE 2008)*, Fortaleza, Ce, nov.

H. Lausen, , Y. Ding , M. Stollberg, D. Fensel, R.L. Hernandez, S. K. Han (2005). Semantic Web Portals: State of the Art Survey. *Journal of Knowledge Management*, 9(5):40-49, maio.

Fernanda Lachtim, Ana Maria de C. Moura, Maria Cláudia Cavalcanti (2009). Ontology Matching for Dynamic Publication in Semantic Portals. Submetido ao *Journal of Brazilian Computer Society (JBACS)*.

J. Madhavan, L. Afanasiev, L. Antova, A. Halevy (2009). Harnessing the deep Web: Present and Future, *4th Bienal Conference on innovative data systems (CIDR)*, Asilomar, CA, EU, jan.

A. Maedche, et al. (2001). “SEmantic portAL - The SEAL Approach”, in *Creating the Semantic Web*, D. Fensel, J. Hendler, H. Lieberman, W. Wahlster (eds.), MIT Press, MA, Cambridge.

A. Maedche, B. Motik, N. Silva and R. Volz (2002). Mafra - A Mapping Framework for Distributed Ontologies. Em *EKAW '02: Proceedings of the 13th International Conf. on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, pages. 235-250, London, UK. Springer-Verlag.

E. Mäkelä, E. Hyvönen, S. Saarela, K. Viljanen (2004). Ontoviews - a Tool for Creating Semantic Web Portals. In *International Semantic Web Conference*, Hiroshima, pages 797-811.

- E. Mäkelä, K. Viljanen, P. Lindgren, M. Laukkanen, E. Hyvönen (2005). View-Based User Interfaces for Information Retrieval on the Semantic Web. In *Proceedings of the ISWC-2005 Workshop End User Semantic Web Interaction*, Nov.
- Ana Maria de C. Moura (2009). CV Lattes: <http://lattes.cnpq.br/4044140234487269>
- M. Nagy, M. Vargas-Vera, E. Motta (2007). DSSim - Managing Uncertainty on the Semantic Web. *Proceedings of the 2nd Ontology Matching Workshop*, Busan, Korea.
- N. F. Noy, e C. D. Hafner (1997). The state of the art in ontology design: A Survey and Comparative Review. *AI Magazine*, 18(3):53-74.
- N. Noy e M. Musen (2002). The Prompt Suite: Interactive Tools for Ontology Merging and Mapping. *Technical report, SMI*, Stanford University, 2002.
- N. Choi, Il-Y. Song, H. Han (2006). Survey on Ontology Mapping. *ACM SIGMOD Record*, 35:34-41.
- W. A. Pinheiro, A.M.C. Moura (2004). An Ontology Based-Approach for Semantic Search in Portals. *Proc of the WEBS 2004*, Saragoza – Spain, set.
- S. Ranganathan e M. Gopinath (1967). Prolegomena to Library Classification. *Asia Publishing House*, New York.
- A. Saldhana (2007). Secure E-Government Portals - Building a web of trust and convenience for global citizens. *W3C Workshop on e-Government and the Web, National Academy of Sciences*, Washington DC, USA.
- T. Sidoroff, E. Hyvönen (2005). Semantic e-Government Portals - a Case Study. *Proceedings of the ISWC-2005 Workshop Semantic Web Case Studies and Best Practices for eBusiness SWCASE05*.
- O. Suominen, K. Viljanen, E. Hyvönen (2007). User-Centric Faceted Search for Semantic Portals. *Proc. of the European Semantic Web Conference ESWC*. Innsbruck, Áustria,.
- A. M. Toledo (2002). Portais corporativos, uma ferramenta estratégica de apoio a gestão do conhecimento. *Dissertação de Mestrado*, COPPE, UFRJ.
- United Nations e-Government Survey (2008) - From e-Government to Connected Governance. Available at <http://unpan1.un.org/intradoc/groups/public/documents/UN/UNPAN028607.pdf>, last access, Fev. 2009.
- O. Valkeapää, O. Alm, E. Hyvönen (2007). An Adaptable Framework for Ontology-based Content Creation on the Semantic Web. *Journal of Universal Computer Science*, vol. 13, no. 12,1835-1853.