



## Portais do conhecimento de universidades: um quadro referencial para avaliação de potencial semântico

*University knowledge gateways: a reference framework for assessment of semantic potential*

Josefina Aparecida Soares Guedes \*

Faimara do Rocio Strauhs \*\*

### RESUMO

Os modelos tradicionais de recuperação da informação utilizados na web são inadequados para uma busca efetiva e rápida, e os resultados são, comumente, numerosos e irrelevantes. Por outro lado, os modelos baseados em ontologias e tecnologias de web semântica oferecem maior rapidez no acesso às informações com relevância dos resultados. O objetivo desta pesquisa foi propor condições mínimas necessárias para o estabelecimento de potencial semântico de portais. Um questionário foi aplicado para coletar dados da equipe técnica responsável pelos portais de universidades públicas federais brasileiras da região Sul do Brasil, a fim de identificar ferramentas e tecnologias utilizadas nos portais. Como principais resultados, apresentam-se um diagnóstico do uso da ontologia na estruturação dos portais universitários pesquisados e a proposição de um quadro de referência das condições mínimas necessárias para estabelecimento de potencial semântico dos portais.

**Palavras-chave:** Recuperação de Informação Online; Ontologia; Web Semântica; Portal; Universidades.

### ABSTRACT

Traditional models of information retrieval used on the web are inadequate for effective and quick search and the results are often numerous and irrelevant. On the other hand, models based on ontologies and semantic web technologies offer greater speed of access to information with relevant results. The goal of this research was to propose minimum necessary conditions for establishment of semantic potential of gateways. A questionnaire was used to collect data from the technical team responsible for homepages of federal public universities in southern Brazil, in order to identify tools and technologies used in these homepages. The main results presented were a diagnosis of ontology use in structuring the investigated university portals and the proposal of a framework containing the minimum necessary conditions to establish their semantic potential.

**Keywords:** Online Information Systems; Management Information Systems; Information Retrieval; Ontology; Universities.

\* Mestre em Tecnologia pela Universidade Tecnológica Federal do Paraná, Brasil. Bibliotecária da Universidade Federal do Paraná. Endereço: Rua Padre Camargo, nº 280, 1º andar, Centro, CEP 80060-240, Curitiba, PR. Telefone: (41) 3360-7278. E-mail: Josefinaguedes@yahoo.com.br.

\*\* Doutora em Engenharia de Produção pela Universidade Federal de Santa Catarina, Brasil. Profa. da Universidade Tecnológica Federal do Paraná, Brasil. Endereço: Av. Sete de Setembro, 3165, Rebouças, CEP 80230-901, Curitiba, PR. Telefone: (41) 3310-4713. E-mail: faimara@utfpr.edu.br

## INTRODUÇÃO

Recurso imprescindível para o sucesso das organizações, o conhecimento é formado por elementos tácitos, advindos das práticas diárias, e por elementos explícitos, registrados cientificamente (POLANYI, 1966). Esses dois elementos são convertidos em informações assimiladas por meio de ações de socialização, de externalização, de combinação e de internalização, e propiciam a formação de redes de conhecimentos (NONAKA; TAKEUCHI, 1997).

Com a difusão da internet, o acesso ao conhecimento tem sido facilitado na *web*. Entretanto, ainda persiste a dificuldade de se extrair informações relevantes diante do acúmulo de fontes de informação disponibilizadas. Tais dificuldades se refletem fortemente na gestão do conhecimento, uma vez que a inovação nas diferentes áreas está diretamente associada ao uso de informações criteriosamente selecionadas (STRAUHS et al., 2012). Nesse contexto, os métodos, as técnicas e as ferramentas de gestão do conhecimento e de gestão da informação podem ser utilizados para aperfeiçoar o uso do conhecimento criado (AL BALUSHI; SAMPAIO; LOUCOPOULOS, 2013; WIMMER; YOON; RADA, 2013).

Os portais corporativos são apontados como instrumentos de gestão do conhecimento, pois utilizam ferramentas e tecnologias de informação e comunicação que oferecem vantagens para integração, organização, colaboração, disseminação e recuperação de informações em diferentes contextos organizacionais (RAUTENBERG; TODESCO; STEIL, 2010). No âmbito educacional, para atingir seus propósitos, os portais universitários devem ser construídos de forma a atender critérios de usabilidade que incluem, além da facilidade de uso, a aprendizagem, a colaboração e a satisfação de seus usuários.

A estruturação de portais baseada nos princípios da gestão do conhecimento e gestão da informação, aliada ao uso de ontologia formal constitui uma alternativa para melhorar a recuperação de informação com maior precisão e rapidez (GUARINO, 1997; ZHOU; DING; FININ, 2011; VITAL, 2012). Adicionalmente, o uso de tecnologias e ferramentas da *web* semântica pode ampliar as possibilidades de pesquisa e a oferta de serviços de informação no portal. Assim, uma abordagem interdisciplinar, envolvendo as áreas da gestão do conhecimento, da gestão da informação, ontologia e *web* semântica, pode contribuir para a avaliação do potencial semântico dos portais.

Este trabalho tem por objetivo propor um quadro referencial com condições mínimas necessárias para o estabelecimento do potencial semântico de portais universitários a partir de diagnóstico em portais de universidades públicas federais da região Sul do Brasil. Descrevem-se sequencialmente, a partir do escopo do artigo, alguns princípios da ontologia formal e de sua aplicação em sistemas de recuperação da informação, delineando-se também conceitos de portais do conhecimento, portais semânticos e suas relações com a *web* semântica, base teórica necessária para a proposição pretendida.

## ONTOLOGIA COMO MODELO DE RECUPERAÇÃO DA INFORMAÇÃO EM PORTAIS DO CONHECIMENTO

O termo ontologia foi conceituado por Aristóteles como sendo “categoria” que pode ser usada para classificar alguma coisa. Na organização da informação, a ontologia pode ser utilizada para especificar uma conceituação compartilhada, formal e explícita, sendo possível a combinação de termos e de relações, incluindo

propriedades, funções, restrições e axiomas, explicitamente definidos (GRUBER, 1993; GUARINO, 1997).

Por se caracterizar uma representação estruturada e formal de um determinado domínio de conhecimento, a ontologia permite que máquinas (agentes de *software*) compreendam a semântica das informações, inclusive inferindo, automaticamente, novos dados, ampliando, portanto, a base de conhecimento dos portais. Além disso, o componente formal possibilita o compartilhamento de ontologias, estabelecendo um vocabulário comum facilitador de integração semiautomática de informações de fontes de dados distintas, facilitando a gestão e a recuperação de informações (NOY; MCGUINNESS, 2005).

Dessa forma, o uso de ontologia formal possibilita a recuperação da informação a partir da combinação de relações entre os termos, o que proporciona maior precisão e rapidez no acesso à informação, com um resultado mais seletivo da informação, quando comparado com o uso de palavras-chave isoladas, convencionalmente utilizadas nas pesquisas em portais do conhecimento estruturados com taxonomias leves ou pouco desenvolvidas (WANG et al., 2011; FERNÁNDEZ et al., 2011; BELLANDI et al., 2012; AL BALUSHI; SAMPAIO; LOUCOPOULOS, 2013; WIMMER; YOON; RADA, 2013).

A aplicação de ontologia aos sistemas de informação e aos portais do conhecimento foi impulsionada a partir do advento da *web* semântica, cujos fundamentos estão delineados a seguir.

## PORTAIS DO CONHECIMENTO E WEB SEMÂNTICA

Os portais do conhecimento consistem em um ponto único de acesso aos recursos informacionais, e têm por objetivo oferecer rapidez, confiabilidade, unicidade e facilidade de acesso à informação distribuída em diversos sistemas, arquivos e base de dados (WIMMER; YOON; RADA, 2013). Para que esses portais propiciem aprendizado, socialização e disseminação do conhecimento, é preciso organizar as informações de modo que estas possam ser acessadas/recuperadas rapidamente. Entretanto, o aumento exponencial das fontes de dados dificulta a seleção, a aquisição e a combinação de dados de interesse e, para que o acesso ao conhecimento seja facilitado, faz-se necessário que as informações estejam estruturadas e classificadas.

A dificuldade na recuperação de informações relevantes nos portais e na *web* cresce na mesma medida em que ocorre o crescimento exponencial e desordenado da informação, sendo iminente uma solução que facilite a busca e melhore a seleção de informações. Nesse contexto, surge a *web* semântica, apresentada por Tim Berners-Lee em 2001, também chamada por alguns autores de *web* 3.0 (BERNERS-LEE; HENDLER; LASSILA, 2001; KURILOVAS; KBILINSKIENE; DAGIENE, 2014).

A proposta da *web* semântica teve por objetivo inserir significado nos documentos para que as máquinas também pudessem entendê-los e, dessa forma, aumentar a eficiência das buscas e facilitar a recuperação de informações relevantes para o usuário da *web* (BERNERS-LEE; HENDLER; LASSILA, 2001).

A *web* semântica apresenta uma estrutura dividida em camadas de linguagens, propostas por Koivunen e Miller (2001). Além das tecnologias de base, que incluem sistema operacional, banco de dados, servidor *web* e sistema de gerenciamento de conteúdo, usadas também na *web* tradicional, a estrutura semântica inclui

tecnologias específicas e necessárias para que os conteúdos das páginas possam ser compreendidos por computadores, ou seja, pela ação de agente inteligente, o *software*, que por meio de uma infraestrutura de dados, utiliza técnicas de inteligência artificial que permitem a troca de informações, automatizando tarefas de rotina dos usuários (BERNERS-LEE; HENDLER; LASSILA, 2001).

Os portais que utilizam ontologia em sua modelagem e incluem tecnologias da *web* semântica são classificados como portais semânticos, que propiciam maior eficiência na recuperação da informação se comparado aos portais tradicionais (LAUSEN et al., 2005; WIMMER; YOON; RADA, 2013).

## **PORTAIS SEMÂNTICOS: ESTRUTURA E CARACTERÍSTICAS**

Os portais semânticos constituem uma extensão natural dos portais do conhecimento tradicionais, porém com características distintas quanto à navegação, atualização, processamento, acesso e publicação de dados e informações (BERNERS-LEE; HENDLER; LASSILA, 2001; REYNOLDS; SHABAJEE; CAYZER, 2004; LAUSEN et al., 2005).

Os portais tradicionais estão estruturados de forma a facilitar a busca por texto livre em categorias predefinidas, com uma hierarquia única de classificação cuja tendência é considerar a informação apenas sob um aspecto, o que induz o usuário a uma visão semelhante sobre um determinado termo (CORRÊA, 2012). Além disso, as páginas não são descritas de forma adequada, sem separação entre os conteúdos dos documentos e a apresentação da informação. A falta de um padrão na apresentação das informações dificulta a recuperação eficiente destas nos portais tradicionais (FERNÁNDEZ et al., 2011; BELLANDI et al., 2012; CORRÊA, 2012).

Já os portais semânticos são caracterizados por oferecerem serviços e troca de informação com base na utilização de tecnologias da *web* semântica incluindo linguagens ontológicas, tais como Web Ontology Language (OWL) e Resource Description Framework Schema (RDFS), que possuem descritores lógicos para prover interoperabilidade semântica e adição de valor semântico para o tratamento da informação. Além disso, é possível extrair conteúdos das instâncias encontradas em outros portais para alimentar de forma dinâmica e automática um portal semântico (LACHTIM; MOURA; CAVALCANTE, 2009).

O uso de ontologia de domínios específicos possibilita uma hierarquia flexível de classificação, com incorporação de homônimos e sinônimos de conceitos que melhoram os resultados obtidos (LACHTIM; MOURA; CAVALCANTI, 2009; GAVRILOVA, 2011; WIMMER; YOON; RADA, 2013). O processamento de informações por máquinas constitui uma marcante característica dos portais semânticos, já que o compartilhamento da representação do conhecimento por agentes inteligentes auxilia os usuários em suas buscas, aumentando a relevância dos resultados (LACHTIM; MOURA; CAVALCANTI, 2009; CORRÊA, 2012).

Uma descrição genérica de uma arquitetura para portais semânticos, a partir de uma intranet, foi apresentada por Maedche et al. (2001), com base na estruturação da *web* semântica. Lausen et al. (2005) também delinearão uma estrutura para portais semânticos composta em três camadas: 1- Camada de acesso a informações – que envolve as características de interação entre usuário e o sistema; 2- Camada de processamento de informações – que abrange o caminho da informação processada pelo portal, e que é composto por cinco passos: (i) criação, (ii) publicação, (iii) organização, (iv) acesso e (v) manutenção. Inclui também recursos de colaboração e;

3- Camada de tecnologias de suporte, que inclui tecnologias de base necessárias para que as demais camadas funcionem adequadamente, quer sejam: tecnologias de *web* semântica, ontologias e serviços de *web* semântica.

De acordo com Reynolds, Shabajee e Cayzer (2004), um portal semântico deve apresentar as seguintes características: 1- busca por ontologia; 2- navegação multidimensional com facetas; 3- adição de novas classificações na estrutura da ontologia; 4- gerenciamento descentralizado de conteúdos; 5- processamento automatizado para acesso de agentes inteligentes; 6- integração automática de ontologias.

Para Corrêa (2012), um portal pode ser considerado potencialmente semântico quando apresenta *links*, listas, tabelas e algum tipo de organização e hierarquia em sua estrutura, ou quando parte de seu conteúdo é apresentado sob a forma de uma taxonomia, bastando que apenas algumas de suas páginas apresentem tais características. Entretanto, para ser considerado semântico no todo, o portal deve incluir todas as camadas integrantes da arquitetura computacional característica da *web* semântica.

## PROCEDIMENTOS METODOLÓGICOS

Este estudo foi predominantemente bibliográfico, mas também com pesquisa de campo do tipo diagnóstico, a partir de múltiplas unidades experimentais, utilizando questionário como instrumento metodológico (GIL, 2010). Para tanto, foi realizado um levantamento bibliométrico, com análise de publicações científicas, revisadas e indexadas em bases de dados de relevância das áreas relacionadas ao estudo. As estratégias de busca, as bases de dados utilizadas e o número de artigos recuperados estão ilustrados no quadro 1. Após a exclusão de artigos duplicados entre as bases e daqueles não alinhados ao tema, o portfólio ficou com 75 documentos que foram analisados em termos de conteúdo e do número de citações, compondo o *corpus* de pesquisa resultante de uma análise sistêmica. Estabeleceu-se como valor de corte documentos que foram citados pelo menos uma vez nas bases de dados internacionais, sendo arrolados 34 documentos que obtiveram ao todo 4.113 citações. A análise destas citações permitiu identificar os autores de base e validar o parâmetro temporal da pesquisa bibliométrica na referida área do conhecimento.

**Quadro 1 – Pesquisa bibliométrica.**

PESQUISA BIBLIOMÉTRICA		
ESTRATÉGIAS DE BUSCA	BASE DE DADOS	Nº DE ARTIGOS RECUPERADOS
Ontology* AND semantic portal AND knowledge portal; Ontology* AND portal construction; Ontology* AND portal construction; Ontology* AND (academic portal OR scholar portal OR university portal); Semantic Web AND (academic portal OR scholar portal OR university portal).	Web of Science – Coleção Básica.	167
	Library & Information Science Abstracts (LISA) e Cumputer And Information Systems Abstracts (CISA).	179
	Scielo.	215
	Banco de Teses e Dissertações.	22

	TOTAL	583
--	-------	-----

Fonte: Aatoria própria (2015).

Do ponto de vista dos procedimentos técnicos, foram utilizados métodos mistos, para a coleta e as análises efetuadas. Conforme Creswell e Clark (2013, p. 28) na pesquisa por métodos mistos, é possível utilizar números e palavras para articulação sequencial de pensamento dedutivo e indutivo, caracterizando, assim, uma pesquisa qualiquantitativa.

A partir da análise de conteúdo da literatura levantada, foi elaborado um questionário, de acordo com as características de portais semânticos, que foi endereçado aos responsáveis pela área de tecnologia da informação das 11 universidades federais da região Sul do Brasil, a fim de identificar as ferramentas e as tecnologias utilizadas, e se estas apresentavam alguma característica de *web* semântica. A estrutura básica do questionário, composto de 28 questões, está ilustrada no quadro 2.

**Quadro 2 – Estrutura do questionário da pesquisa.**

<b>ESTRUTURA DO QUESTIONÁRIO DA PESQUISA</b>		
<b>QUESTÕES</b>	<b>DETALHAMENTO</b>	<b>Nº. DE QUESTÕES</b>
<b>ELENCAR CONDIÇÕES PARA ESTABELECEER POTENCIAL SEMÂNTICO</b>	Uso de sistemas abertos.	3
	Formas de armazenamento de dados.	1
	Formas de gestão da informação.	8
	Gestão de ontologia.	5
<b>AVALIAR O PORTAL</b>	Acesso à informação e tecnologia usada.	11

Fonte: Guedes (2015).

Para identificar o potencial semântico dos portais, foram formuladas questões que abordaram: sistema operacional, banco de dados e servidor *web* utilizados; sistema de repositório de documentos e navegação pelos dados; sistema de recuperação da informação; sistema gerenciador de conteúdo e a forma de gestão do conteúdo; edição, atualização e forma de processamento de itens de informação; tecnologias para importação e exportação de dados; mecanismos de inferência e linguagens de programação. Para avaliação do portal na percepção dos respondentes, foram incluídas questões associadas à criação, publicação, organização, acesso e manutenção da informação, e recursos de colaboração disponíveis. O quadro 3 destaca o tratamento dos dados coletados.

**Quadro 3 – Tratamento dos dados coletados.**

TRATAMENTO DOS DADOS COLETADOS		
GRUPO DE QUESTÕES	NO. DE QUESTÕES	TIPO DE ANÁLISE
Identificação das condições para estabelecer potencial semântico.	17	Análise descritiva percentual.
Avaliação do portal na percepção dos respondentes.	11	Análise estatística descritiva.

Fonte: Guedes (2015).

As variáveis quantitativas, utilizadas no questionário, foram medidas pela escala de Likert (1 a 5) e a representatividade das médias ( $\bar{x}$ ) foi registrada de acordo com os seguintes intervalos: 1 a 2,9 – distante do ideal em relação às características semânticas; 3 a 3,9 – no limiar; ligeiramente próximo ou mais próximo em relação às características semânticas; 4 a 4,9 – muito próximo ao ideal em relação às características semânticas e; 5 – ideal característico de estruturação semântica, definido em termos do uso de tecnologias e ferramentas da web semântica e da ontologia como modelo de recuperação da informação. Foram analisados ainda o desvio padrão (DP) e o coeficiente de variação (CV), sendo este considerado representativo do conjunto de dados quando o valor for menor que 0,5, ao passo que valores acima deste patamar indicam que a média não foi representativa, ou seja, ela está sendo influenciada por valores extremos (TRIOLA, 1999).

A análise quantitativa e qualitativa foi complementada por uma observação sistemática, que, segundo Marconi e Lakatos (2010, p. 76-78), é “o ponto de partida da investigação social”, e que pode se valer de quadros como instrumento de coleta. Segundo Gil (2010, p. 121), a observação sistemática auxilia na parte descritiva da pesquisa, mas também como um elemento exploratório, familiarizando pesquisadores com o seu objeto de pesquisa.

Neste estudo, a observação sistemática foi operacionalizada mediante uma pesquisa efetuada nos portais gerais das universidades pesquisadas, a partir de uma mesma palavra-chave, “calendário acadêmico”, com o intuito de coletar dados referentes à estruturação e à organização da informação, que podem ser identificadas pelo usuário final. A análise estatística envolveu variáveis de sequência que identificaram as tecnologias e as ferramentas utilizadas nos portais e variáveis quantitativas que registraram a percepção dos respondentes sobre o acesso à informação dos portais. As análises foram tratadas com os aplicativos Statistical Package for the Social Sciences e Microsoft Excel.

## RESULTADOS E DISCUSSÃO

Na região Sul do Brasil, localizam-se 11 universidades federais, das quais 9 responderam ao questionário. A partir da análise das variáveis de sequência e da observação sistêmica, foi possível verificar que na maioria das universidades os portais são gerenciados pela equipe do setor de tecnologia da informação (73%), e que em duas instituições a gestão é feita conjuntamente pelas áreas de comunicação e de TI (18%). O trabalho interdisciplinar está em concordância com os preceitos da web semântica e da modelagem computacional ontológica, pautados em ambientes

colaborativos e interdisciplinares que integram pessoas e processos envolvidos com a informação e conhecimento (FERNÁNDEZ et al., 2011; WINNER; YOON; RADA, 2013).

Quanto às tecnologias de base, verificou-se que o sistema operacional utilizado em todos os portais foi o Linux. O banco de dados usado pela maioria foi o MySQL (67%), e para o servidor *web*, a maioria (89%) utilizou uma especificação do Apache. Quanto ao gerenciador de conteúdo, o destaque ficou para o WordPress (45%), Joomla (11%) e Drupal (11%). Todas essas tecnologias são compatíveis com sistemas de código e licença aberta, recomendados para intercâmbio e reuso de dados. Além disso, tais tecnologias possibilitam o intercâmbio e o reuso de dados em diferentes aplicativos, o acesso democrático à informação e ao conhecimento novo a partir da combinação de fontes de dados e padrões (BERNERS-LEE; HENDLER; LASSILA, 2001; MAEDCHE; STAAB, 2001; LAUSEN et al., 2005; WORLD WIDE WEB, 2014).

Quanto à gestão da informação e de ontologia, verificou-se que nenhum portal utiliza ontologia em sua estruturação. Essa condição foi ratificada pela observação sistêmica, na qual se verificou que a opção para pesquisa na maioria dos portais (89%) foi o uso de palavras-chave, que está sujeito à falta de precisão nos resultados, quando comparado ao uso de ontologia, que fornece resultados mais seletivos e relevantes (FERNÁNDEZ et al., 2011; BELLANDI et al., 2012; WINNER; YOON; RADA, 2013; AL BALUSHI; SAMPAIO; LOUCOPOULOS, 2013). Isso porque, embora os motores de buscas classifiquem termos de pesquisa baseado em taxonomias, em muitos casos, devido à polissemia e à sinonímia, os termos podem estar relacionados a conceitos diferentes daqueles que o usuário busca (BARROS; GONÇALVES; SANTOS, 1998).

A maioria dos portais não dispõe de mecanismos de inferência (67%) e as tecnologias citadas foram SPARQL (11%), SQL (11%) e XQuery (11%) que figuraram em apenas um portal, cada uma isoladamente. Tecnologias de importação e exportação (89%) não foram observadas na maioria dos portais e, também na maior parte destes, o processamento e a publicação de informação é feito de forma manual (89%). Tais condições dificultam a integração entre sistemas e o acesso de agentes de *software* inteligentes, inibindo também o compartilhamento de informação e a combinação de raciocínios automatizados para melhorar a qualidade da informação recuperada, tal como ocorre em portais semânticos (BERNERS-LEE; HENDLER; LASSILA, 2001; KOIVUNEN; MILLER, 2001; MAEDCHE STAAB, 2001; MAEDCHE et al., 2001; FERNÁNDEZ et al., 2011; GAVRILOVA, 2011; BELLANDI et al., 2012).

A gestão de conteúdo em boa parte dos portais é feita de forma centralizada (56%) e o armazenamento de dados é feito, majoritariamente, mediante classificação hierárquica fixa em processo manual (89%). Entretanto, a descentralização de conteúdo e a navegação multidimensional com facetas e arranjo de informações semiestruturadas e extensíveis são características mais compatíveis com portais semânticos em que há possibilidade de múltiplas agregações e visões do mesmo dado (REYNOLDS; SHABAJEE; CAYZER, 2004; LAUSEN et al., 2005; CORRÊA, 2012).

A linguagem de marcação usada por todos os portais foi a HyperText Markup Language (HTML). Apesar da inclusão de recursos que podem fornecer notação semântica em versões mais atuais da HTML, esta tecnologia tem sido caracterizada mais pela marcação sintática de texto, e não pela marcação semântica, que pode ser interpretada pelos mecanismos de inferência da *web* semântica (ABANDA et al., 2013). Tal condição foi ratificada na observação sistêmica e denota que os portais avaliados, neste quesito, estão mais distantes do modelo semântico de recuperação da informação. De acordo com os valores encontrados

**Tabela 1 – Valores das variáveis quantitativas analisadas no questionário.**

VARIÁVEIS ANALISADAS	MÉDIA	DESVIO PADRÃO	COEFICIENTE DE VARIAÇÃO
Maturidade de implantação do portal.	4,11	0,6	0,14
Informações relevantes do domínio da comunidade acadêmica.	4,22	1,30	0,30
Profundidade de cobertura de informações com relevância semântica dos recursos de informação do portal.	3,33	0,5	0,15
Disponibilização de informação semanticamente completa e consistente.	3,22	0,83	0,25
Utilização de vocabulário e ações sem ambiguidade.	3,55	0,72	0,20
Presença de mecanismos de ajuda ou mapa do site.	3,00	0,70	0,23
Presença de materiais de consulta para descrição das rotinas de processamento semântico ou estrutura da ontologia	1,55	0,88	0,56
Disponibilização de recursos de colaboração, comunicação e compartilhamento de informação.	2,00	0,86	0,43
Presença de mapas de tópicos pessoais para melhorar a gestão de informação pessoal e a eficácia do portal	2,22	1,2	0,54
Uso de recurso de marcação semântica para facilitar a troca de informação.	2,88	0,78	0,27
Personalização de funcionalidades do portal.	2,66	1,32	0,49

Fonte: Guedes (2015).

para as variáveis quantitativas utilizadas no questionário, listadas na tabela 1, as médias das variáveis de maturidade de implantação do portal, sem *links* quebrados e de informações relevantes do domínio da comunidade acadêmica foram consideradas próximas ao ideal semântico, pois retratam a atualização e a profundidade de cobertura de informações que atendem aos princípios da *web* semântica.

Comparando a análise das variáveis de sequência com a observação sistêmica, verificam-se algumas discrepâncias. Na maioria dos portais (64%), os resultados da busca pela palavra-chave “calendário acadêmico” foram exibidos em *links* que continham os dois termos juntos ou isolados, ordenados aleatoriamente, denotando a recuperação por marcação simples das palavras-chave no título de documentos ou notícias, e não por classificação e recuperação semântica do conteúdo possível a partir do uso de taxonomia ou ontologia. Em apenas um portal (9%), os *links* foram ordenados alfabeticamente pela palavra-chave, indicando uma maior organização dos resultados exibidos. Do mesmo modo, apenas em um portal foi indicado o uso de taxonomia, embora não tenha sido possível confirmar tal informação, podendo tratar-se apenas de uma simples classificação sintática baseada no número de ocorrências das palavras. Entretanto, a observação sistêmica indicou que dois portais

(18%) apresentaram uma categorização prévia de palavras-chave associadas ao assunto pesquisado para seleção no momento da pesquisa. Tal estruturação aproxima-se do arranjo taxonômico característico de portais com algum potencial de marcação semântica.

Ficaram no limiar em relação ao modelo semântico as variáveis: profundidade de cobertura de informações com relevância semântica dos recursos de informação do portal; disponibilização de informação semanticamente completa e consistente; utilização de vocabulário e ações sem ambiguidade; e presença de mecanismos de ajuda ou mapa do site. Tais resultados denotam a presença de algumas características semânticas nos portais que podem favorecer a transição de um modelo tradicional de estruturação para o modelo semântico de recuperação da informação.

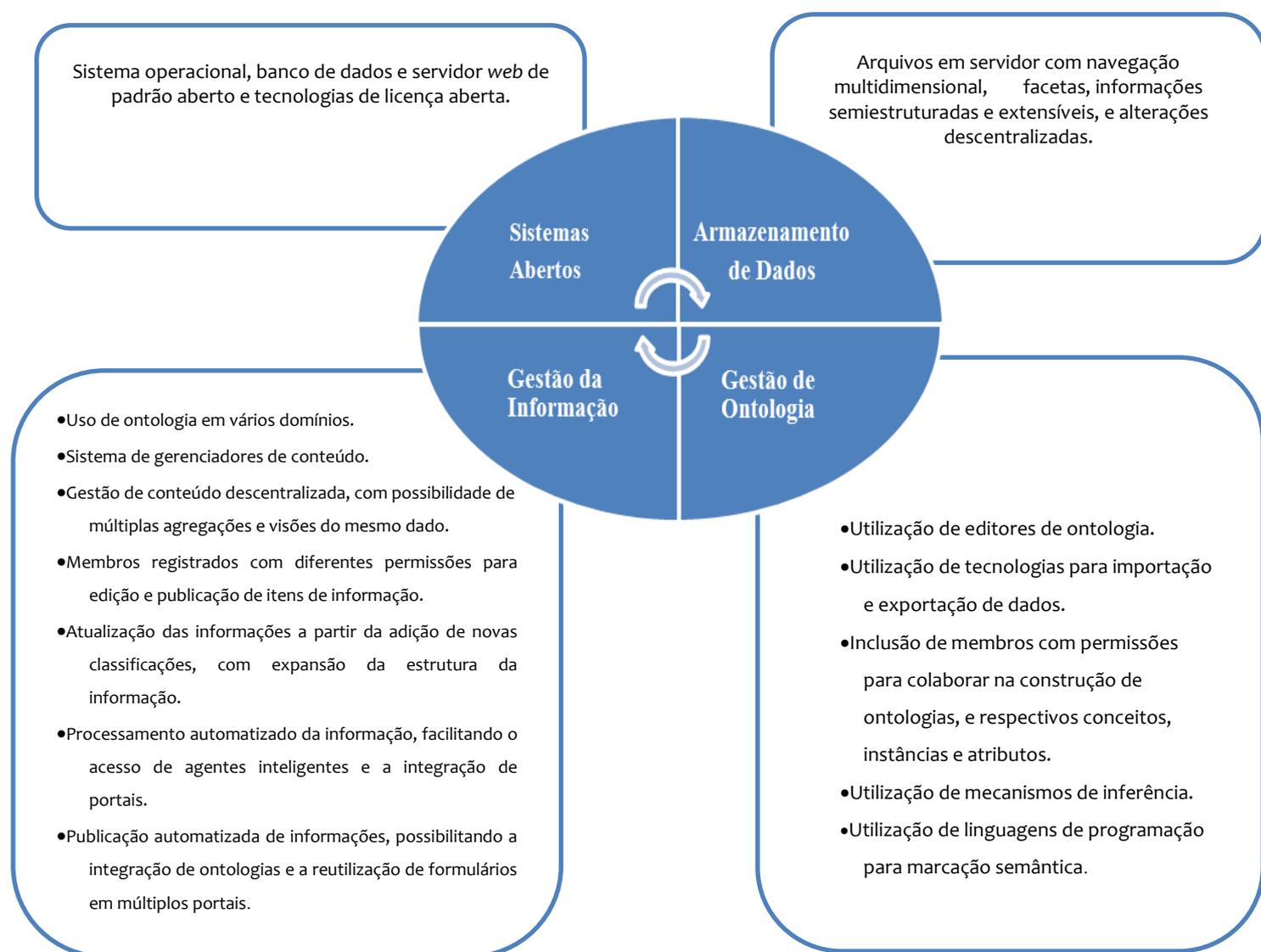
Distantes do ideal semântico de estruturação dos portais, figuraram as variáveis: presença de materiais de consulta para descrição das rotinas de processamento semântico ou estrutura da ontologia; disponibilização de recursos de colaboração, comunicação e compartilhamento de informação; presença de mapas de tópicos pessoais para melhorar a gestão de informação pessoal e a eficácia do portal; uso de recurso de marcação semântica para facilitar a troca de informação; e personalização de funcionalidades do portal. Na observação sistêmica, também se verificou o uso de algum recurso de informação nos portais (54%), tais como a opção com filtros dos resultados por relevância e data. Os serviços de personalização traduzem-se em facilidades que permitem aos usuários organizarem e manipularem a informação obtida, constituindo, assim, atrativos dos portais (GAVRILOVA, 2011; BELLANDI et al., 2012; AL BALUSHI; SAMPAIO; LOUCOPOULOS, 2013).

Na análise da distribuição das médias obtidas para as variáveis quantitativas, observou-se um baixo percentual de médias, 18%, que se aproximou do modelo ideal semântico de portal. Grande parte dos portais, 45%, distanciou-se do ideal, e 36% dos portais ficaram no limiar da escala, denotando a presença de algumas características semânticas de estruturação.

Analisando as médias encontradas para as questões de sequência relacionadas às tecnologias usadas e às formas de gestão, também se nota que apenas 29% dos portais utilizam tecnologias características de modelo semântico, tais como uso de tecnologias de padrão aberto e publicação e edição de informações por membros registrados com diferentes permissões. Por outro lado, 71% dos portais apresentam características associadas ao modelo tradicional de recuperação da informação.

Correlacionando os resultados encontrados e a literatura, apresenta-se um quadro referencial com as condições mínimas para estabelecimento de potencial semântico de portais, estruturadas a partir de quatro tópicos que sinalizam o uso de sistemas de licença aberta, a forma de armazenamento de dados, e estabelecem características para gestão da informação e gestão de ontologia (Quadro 4).

#### Quadro 4 – Quadro referencial de condições mínimas para estabelecimento de potencial semântico de portais.



Fonte: Guedes (2015).

### CONSIDERAÇÕES FINAIS

A partir da literatura consultada e dos resultados apresentados, foi possível verificar que o uso de tecnologias e sistemas de padrão aberto e de licença aberta constitui condições fundamentais na estruturação de portais semânticos, pois potencializam o intercâmbio e o reuso de dados, e democratizam o acesso às informações.

A possibilidade de acesso aos dados a partir de navegação multidimensional denota que este tipo de armazenamento de dados favorece o acesso às informações no portal.

Quanto às rotinas de administração e à organização do portal, na gestão de informação, o uso de ontologia é uma condição determinante em portais semânticos. Os gerenciadores de conteúdo, o processamento e a publicação automática, aliados a rotinas de atualização de informações traduzem-se em recursos de interação entre sistema e usuário, que proporcionam o acesso ao portal como uma aplicação de web semântica. Já as rotinas de gestão de ontologia identificam os recursos de processamento e de colaboração, com análise das tecnologias de web semântica utilizadas, dos serviços ofertados e demais tecnologias que possibilitam a construção da web semântica.

A análise dos portais avaliados denotou que nenhum portal incluiu a ontologia em sua estruturação, e, na maioria deles (89%), a recuperação da informação é estruturada pela simples marcação de palavras-chave no título dos documentos. Apenas em 29% dos portais, foram observadas características compatíveis com modelos ontológicos de estruturação. Verificou-se, ainda, que nenhum portal atingiu o patamar ideal de estruturação semântica, e que apenas 18% deles se aproximaram do modelo ideal de estruturação semântica.

Conclui-se que os portais universitários pesquisados utilizam algumas tecnologias de base compatíveis com a web semântica, porém não utilizam a ontologia em sua estruturação, o que os aproxima do modelo tradicional de recuperação da informação. Por outro lado, algumas características de organização e apresentação de informações dos portais sinalizam que estes podem apresentar algum potencial semântico se forem adotadas outras tecnologias propostas no quadro referencial apresentado neste estudo.

Considerando que o conhecimento evolui incessantemente, este trabalho não teve a pretensão de esgotar o assunto e, portanto, recomenda-se a condução de outros estudos para ratificar ou complementar os achados a partir de pesquisas em outras organizações, e que possibilitem a elaboração de um índice contendo parâmetros para a avaliação do potencial semântico de portais. Uma ferramenta nesse escopo pode ser de grande valia na gestão de portais voltados para ações de criação, compartilhamento e difusão do conhecimento dentro ou fora do âmbito acadêmico.

Artigo recebido em 15/09/2015 e aprovado em 15/03/2016.

## REFERÊNCIAS

ABANDA, F. H.; TAH, J. H. M.; KEIVANI, R. Trends in built environment semantic web applications: Where are we today? *Expert Systems with Applications*, v. 40, n. 14, p. 5.563-5.577, 2013.

AL BALUSHI, T. H.; SAMPAIO, P. R. F.; LOUCOPOULOS, P. Eliciting and prioritizing quality requirements supported by ontologies: a case study using the ElicitO framework and tool. *Expert Systems*, v. 30, n. 2, p. 129-151, May 2013.

BARROS, F. A.; GONÇALVES, P. F.; SANTOS, T. L.V. L. Providing context to web searches: the use of ontologies to enhance search engine's accuracy. *Journal of the Brazilian Computer Society*, Campinas, v. 5, n. 2, nov. 1998.

BELLANDI, A. et al. Assisted knowledge base generation, management and competence retrieval. *International Journal of Software Engineering and Knowledge Engineering*, v. 22, n.8, p.1.007-1.037, 2012.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, n. 1, May 2001.

CORRÊA, D. A. *Uma abordagem para extração de conteúdos baseada em características estruturais e navegacionais de portais web*. 2012. 110 fl. Dissertação (Mestrado em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro, 2012.

CRESWELL, J. W.; CLARK, V. L. *Pesquisa de métodos mistos*. 2. ed. Porto Alegre: Penso, 2013.

FERNÁNDEZ, M. et al. Semantically enhanced information retrieval: an ontology-based approach. *Web Semantics: Science and Agents on the World Wide Web*, v. 9, p.434-452, 2011.

GAVRILOVA, E. A. Implementation features a university data portal using the semantic web technology. *Programming and Computer Software*, v. 37, n. 1, p. 48-55, Jan. 2011.

GIL, A. C. *Como elaborar projetos de pesquisa*. 5. ed. São Paulo: Atlas, 2010.

GRUBER, T. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, v. 43, p. 907-928, 1993.

GUARINO, N. Understanding, building and using ontologies. *International Journal of Human and Computer Studies*, v. 45, n. 2/3, p. 2, 1997.

GUEDES, J. A. S. *Portais do conhecimento de universidades: proposta de um quadro referencial para estabelecimento do potencial semântico*. Curitiba, 2015. 169 f. Dissertação (Mestrado em Tecnologia) – Programa de Pós-Graduação em Tecnologia, Universidade Tecnológica Federal do Paraná.

KOIVUNEN, M.; MILLER, E. W3C: semantic web activity. In: SEMANTIC WEB KICK-OFF SEMINAR IN FINLAND, Helsinki, 2001. *Proceedings... Helsinki: Helsinki Institute for Information Technology, 2001*.

KURILOVAS, E.; KBILINSKIENE, S.; DAGIENE, V. Web 3.0: based personalization of learning objects in virtual learning environments. *Computes in Human Behavior*, v. 30, p. 654-662, 2014.

LACHTIM, F. A.; MOURA, A. M. C.; CAVALCANTI, M. C. Ontology matching for dynamic publication in semantic portals. *Journal of the Brazilian Computer Society*, v.15, n.1, p. 27-43, 2009.

LAUSEN, H. et al. Semantic web portals – state of the art survey. *Journal of Knowledge Management*, v. 9, n. 5, p. 40-49, 2005.

MAEDCHE, A. et al. Seal: a framework for developing semantic portals. In: BRITISH NATIONAL CONFERENCE ON DATABASES, 18., Chilton, 2001. *Proceedings... Berlin: Springer, 2001*.

MAEDCHE, A.; STAAB, S. Ontology learning for the semantic web. *IEEE Intelligent Systems & their Applications*, v. 16, n. 2, p. 72-79, Mar.-Apr. 2001.

MARCONI, M. A.; LAKATOS, E. M. *Fundamentos e metodologia científica*. 7. ed. São Paulo: Atlas, 2010.

NONAKA, I.; TAKEUCHI, H. *Criação de conhecimento na empresa: como as empresas japonesas geram a dinâmica da inovação*. Rio de Janeiro: Campus, 1997.

- NOY, N. F.; MCGUINNESS, D. L. *Desarrollo de ontologías – 101: guía para crear tu primera ontología*. 2005. Disponível em: < <http://ocw.uc3m.es/ingenieria-informatica/sistemas-avanzados-de-recuperacion-de-informacion/ejercicios/ontology101-es.pdf/view>>. Acesso em: 19 maio 2013.
- POLANYI, M. *The tacit dimension*. New York: Doubleday, 1966.
- RAUTENBERG, S.; TODESCO, J. L.; STEIL, A. V. Ontologias de domínio no mapeamento de instrumentos da gestão do conhecimento e de agentes computacionais da engenharia do conhecimento: o estado da arte. *Perspectivas em Ciência da Informação*, v. 15, n. 2, p. 163-182, maio/ago. 2010.
- REYNOLDS, D.; SHABAJEE, P.; CAYZER, S. *Semantic information portals*. New York: ACM, 2004. p. 290-291.
- STRAUHS, F. R. et al. *Gestão do conhecimento nas organizações*. Curitiba: Aymar, 2012.
- TRIOLA, Mario F. *Introdução à estatística*. 7. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1999.
- VITAL, L. P. Taxonomia como elemento estruturante em portais corporativos. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 9, n. 2, p. 25-40, jan./jun. 2012.
- WANG, S. et al. A reference ontology based approach for service oriented semantic interoperability. *International Journal of Web Portals*, v. 3, n. 1, 1-16, 2011.
- WIMMER, H.; YOON, V.; RADA, R. Integrating knowledge sources: an ontological approach. *International Journal of Knowledge Management*, v. 9, n. 1, p. 60-75, 2013.
- WORLD WIDE WEB – W3C. *Standards*. Disponível em: <<http://www.w3.org/standards/>>. Acesso em: 13 mar. 2014.
- ZHOU, L.; DING, L.; FININ, T. How is the semantic web evolving? A dynamic social network perspective. *Computers in Human Behavior*, v. 27, n. 4, p. 1.294-1.302, 2011.