XVI SEMEAD
Seminários em Administração
outubro de 2013
ISSN 2177-3866

Seleção de notícias online para inteligência competitiva - Uso de ontologia de domínio do negócio para expansão semântica da busca na internet

CLEBER MARCHETTI DURANTI
USP - Universidade de São Paulo
clebermd@uol.com.br

FERNANDO CARVALHO DE ALMEIDA
USP - Universidade de São Paulo
fcalmeida@usp.br

Seleção de notícias online para inteligência competitiva - Uso de ontologia de domínio do negócio para expansão semântica da busca na internet

1 INTRODUÇÃO

Escanear o ambiente externo e desenvolver interpretações que levem a conhecê-lo são atividades organizacionais importantes e vistas por alguns como requisito básico para a sobrevivência da empresa (DAFT; WEICK, 1986). Espera-se que a empresa que sabe mais a respeito de seus clientes, produtos, tecnologias, mercados e suas conexões, tenha melhor desempenho (ZACK; HALL, 1999).

A internet se apresenta como um recurso externo rico em informações acerca do ambiente e é utilizada extensivamente pelas organizações (MARSHALL *et al.*, 2004). Os autores apontam, porém, a dificuldade em localizar conhecimento novo no vasto volume de informações disponíveis *online*. Trata-se do problema da sobrecarga de informações, a dificuldade em se localizar informação útil em meio ao grande volume de informações irrelevantes (CHUNG; CHEN; NUNAMAKER JR. *et al.*, 2005), fenômeno experimentado, por exemplo, quando se procura informação sobre determinado tema na internet através de um buscador comum e este retorna uma enorme lista de resultados. Trata-se de um problema em aberto para as empresas que utilizam a internet como fonte importante de informações (DAVIS, 2011; DENTON; RICHARDSON, 2012; JENKIN, 2008; LI, 2011; LI *et al.*, 2012; TATE, 2008).

Em sua extensa revisão da literatura acerca de sobrecarga de informações, Eppler e Mengis (2004) descrevem que a sobrecarga ocorre quando o requisito de processamento de informações excede a capacidade de processamento do indivíduo ou da organização. Processamento entendido como coletar, interpretar e sintetizar as informações no contexto das tomadas de decisão da organização.

Davis (2011) descreve a sobrecarga de informações como consequência de abundância de informações e deficiência nos filtros aplicados, e defende que a questão pode ser endereçada pela área de arquitetura de informações. À medida que mais informações se tornam disponíveis na internet, usuários se tornam mais ávidos por ferramentas que os ajudem a filtrar o fluxo de informações e encontrar artigos de seu interesse (MAES, 1994). Eppler e Mengis (2004) sugerem que haverá uma solução definitiva para a sobrecarga de informações, mas que haverá sempre ciclos de refinamentos e melhorias nas buscas de soluções.

Compreender e manter-se atualizado acerca do ambiente externo em que a empresa se insere envolve a descoberta de conhecimentos através de processos de aprendizagem individual e organizacional (JENKIN, 2008). Dado que as pessoas têm capacidade limitada para assimilar novas informações, elas constroem significados atentando seletivamente para informações que se conectam com o que elas já sabem (KUHLTHAU, 1991). Cohen e Levinthal (1990) argumentam que a aprendizagem de novos conceitos deve ser fundada em conhecimento familiar e modelos mentais.

Modelos mentais são estruturas que ajudam a simplificar e organizar as informações (CROSSAN *et al.*, 1999). Referem-se a estruturas de conhecimento que representam o conhecimento como uma rede de conceitos com atributos abstratos, valores, relacionamentos e regras. Tanto indivíduos como organizações têm modelos mentais. No caso das organizações, o modelo mental corresponde a um entendimento compartilhado e negociado. Em aprendizagem organizacional, diversos modos de aprendizagem têm sido discutidos, sendo que alguns deles, como *adjustive learning*, *turnover learning* e *adaptation* tratam de mudanças incrementais e refinamentos em modelos mentais já existentes (JENKIN, 2008).

Em ciência da informação, uma ontologia expressa o conhecimento acerca de um domínio – os conceitos que fazem parte do domínio são representados como nós de uma rede e as relações entre os conceitos são representadas pelos arcos que unem os conceitos, informando-se no arco o tipo de relação existente. Jonker *et al.* (2010) e Kudryavtsev (2006) propõem

ontologias como forma de representação de modelos mentais compartilhados. Ontologias, entendidas como uma representação consensual de um sistema ou organização, ou, como definidas por Gruber (1995), "[...] uma especificação explícita de uma conceitualização." e por Studer *et al.* (1998): "uma especificação explícita formal de uma conceitualização compartilhada de um domínio de interesse". Em outras palavras, uma ontologia descreve, em formato passível de ser processado por uma máquina, um conhecimento comum a um grupo acerca de determinado domínio, definindo seus conceitos, propriedades e atributos em um vocabulário comum ao grupo.

Para Hwang e Salvendy (2005), uma ontologia pode assumir um papel crucial em tornar explícitos modelos mentais individuais e estabelecer modelos mentais compartilhados numa organização. Dentro deste conceito, a existência de uma representação explícita do ambiente competitivo na forma de uma ontologia pode subsidiar a obtenção de informações acerca deste ambiente que leve a incrementos ou atualizações da visão corrente do mesmo.

Este trabalho trata da construção de um sistema de apoio à busca e seleção de informações na internet com base em ontologia representativa do domínio do negócio da empresa, através da expansão semântica dos termos de busca definidos pelo usuário quando da busca de notícias *online* em buscadores-padrão como Google. A expansão visa a adicionar palavras na busca iniciada pelo usuário que reforcem o contexto da informação, melhorando a qualidade dos resultados trazidos pela busca. O sistema aumenta as chances de localização de informações relevantes para o tema em foco, em meio à sobrecarga de informações. O viés introduzido pela ontologia no processo de busca de informação na internet por outro lado, espera-se, tenha efeitos positivos também por estar em linha com os processos de aprendizagem incremental a partir da visão corrente representada pela ontologia. As relações ontológicas podem ainda ser programadas de forma a registrar associações emergentes entre temas, favorecendo a monitoração destas associações.

O sistema de busca apresentado permite maior seletividade que a obtida numa busca padrão pela limitação do volume de notícias trazidas em função da "desambiguação" dos temas, mediante a adição de facetas obtidas das relações ontológicas que qualificam e conferem escopo à informação (GARSHOL, 2004). O sistema colabora ainda para a obtenção de ranqueamento mais apropriado das notícias trazidas, por reforçar o contexto expresso pelos termos adicionados à busca. Como parte do sistema, foi construído um piloto de ontologia sobre o domínio da área de "*outsourcing* de TI" com o qual o sistema foi testado..

O sistema utiliza relações ontológicas para sugerir a adição de termos à busca original. Uma interface interativa permite ao usuário escolher entre as adições sugeridas, de forma a dirimir ambiguidades antes da execução da busca.

1.1 Problema de pesquisa e objetivo

Exploração da aplicação de ontologia de domínio de área de negócio para aumentar a seletividade nas buscas de informações acerca do ambiente competitivo do negócio.

1.1.1 Objetivos específicos

a) Construção de piloto de ontologia do domínio "*outsourcing* de TI";
b) Construção de um sistema de apoio à busca na internet que faça uso das relações entre os conceitos da ontologia para expansão semântica das palavras da busca;
c) Avaliação utilizando o modelo *Technology Acceptance Model* (TAM3).

2 METODOLOGIA *DESIGN RESEARCH*

Conforme Vaishnavi e Kuechler (2004), *Design Research*, ou *Design Science Research*, trata do aprendizado através da construção de artefatos – o "design" (construção de artefato) é utilizado como método ou técnica de pesquisa. Envolve o projeto de novos artefatos e a análise do uso e/ou desempenho dos mesmos para aprimorar e compreender o comportamento

de aspectos de Sistemas de Informação. Tais artefatos incluem, sem se limitar a estes, algoritmos, interfaces homem-máquina e metodologias ou linguagens.
Este trabalho aplica o método *Design Science Research* na construção de dois artefatos: ontologia e sistema para expansão de *query* com base na ontologia, como contramedida para a sobrecarga de informações na busca de notícias na internet.

2.1 Modelo de previsão de aceitação de tecnologia

O TAM – *Technology Acceptance Model* (DAVIS, 1989) foi desenvolvido para a predição individual da adoção e uso de novas tecnologias de TI. Propõe que a intenção individual de usar uma tecnologia é determinada por duas crenças: utilidade percebida – a extensão com que uma pessoa acredita que usar uma tecnologia irá melhorar seu desempenho no trabalho – e facilidade de uso percebida – o grau com que uma pessoa acredita que usar uma tecnologia será livre de esforço. Teoriza ainda que o efeito de variáveis externas, tais como as características do projeto, serão mediadas pela utilidade e facilidade de uso percebidos.. Mais recentemente (VENKATESH; BALA, 2008), o modelo foi expandido, resultando no modelo TAM3 que tem sido capaz de explicar entre 40 e 53% da intenção de uso de tecnologia e, por isto, foi escolhido como base para o teste de aceitação do *software* deste trabalho.
O TAM3 foi adaptado para a avaliação do protótipo tendo em vista de que se trata de uma prova de conceito, não havendo uma situação de introdução real do *software* num ambiente de trabalho, mas um teste simulado por usuários solicitados a experimentar a ferramenta. O conjunto de questões utilizadas pode ser visto na tabela 2 1 (item 4.6.2), onde os usuários indicaram na escala de 1 a 7 o grau de concordância com cada uma dos itens.

3 REVISÃO BIBLIOGRÁFICA

3.1 Sobrecarga de informações na internet

Para Brennan (2006), sobrecarga de informações é ter mais informação do que se consegue adquirir, processar, armazenar ou resgatar. Na revisão de literatura sobre o tema, Eppler e Mengis (2004) dão a seguinte definição: “sobrecarga de informações ocorre quando o suprimento excede a capacidade de consumo”. Para Ong *et al.* (2005), a sobrecarga de informação é decorrência da possibilidade de capturar e acessar um grande volume dados e bases de conhecimento propiciada pela tecnologia da informação. Segundo Davis (2011), o problema não está na abundância de informação, mas na falha em filtrar a informação quando ela é publicada ou consumida e, com a facilidade e o baixo custo de publicação na internet, o filtro de qualidade passou a ficar mais à jusante no processo. Village (2000) entende os motores de busca como a primeira tentativa para se lidar com a sobrecarga de informações na web e os vê como primitivos atualmente.
Conforme Lankton *et al.* (2012), numa busca com base em palavras-chave, boa parte da informação obtida é irrelevante para as necessidades do usuário. A inconsistência entre o que se está procurando e o resultado da busca leva potencialmente à percepção de sobrecarga de informações. Para os autores, uma ferramenta de busca de informação que proveja um direcionamento aos usuários à medida que eles exercitam a busca deve produzir melhores resultados de busca e melhorar o desempenho das ferramentas de busca, mesmo que incrementalmente, podendo trazer benefícios importantes tendo em vista o uso massivo que se faz delas atualmente.
Para Wu *et al.* (2006), a internet é a fonte de maior crescimento para coleta de informações de inteligência, e a sobrecarga de informações apresenta um enorme desafio para os analistas que precisam extrair conteúdo.
Kuhlthau (1991) vê a busca de informação como um processo de criação de sentido no qual a pessoa está formando um ponto de vista – o indivíduo está ativamente envolvido em encontrar

significado que se encaixe no que ele já sabe, na sua estrutura de referência. A informação de várias fontes é assimilada dentro do que já é sabido, por meio de uma série de escolhas.
Jenkin (2008) propõe o uso de teorias de aprendizado organizacional na construção de ferramentas para descoberta de conhecimento na internet. Ferramentas que incorporem o modelo mental compartilhado entre os indivíduos na organização podem suportar o aprendizado incremental, fundado no conhecimento já existente. A autora defende que a aprendizagem pode ser suportada por ferramentas e agrupa os modos de aprendizagem conforme sua interação com os modelos mentais e, para cada modo de aprendizagem, propõe ferramentas de suporte:

a) Manutenção do modelo mental – confirmação e validação dos modelos mentais existentes - é suportado pelos motores de busca comum, como o Google;
b) Refinamento do modelo mental – refinamento e ajuste dos modelos mentais existentes, tornando-os mais eficientes e especializados - pode se beneficiar de ferramentas baseadas em ontologias, bem como de tecnologias de web semântica;
c) Construção de modelos mentais – criação de novos modelos, envolvendo mudanças súbitas e reestruturação do conhecimento. - pode ser apoiado por ferramentas de mineração de texto que explicitam conexões entre temas até então desconhecidas.

Pela proposta de Jenkin (2008), ferramentas que incorporem o modelo mental de indivíduos ou da organização, na forma de ontologias ou outras tecnologias de web semântica, podem guiar a aquisição de conhecimento particularmente no segundo modo de aprendizagem visto acima, ao suportar a descoberta de múltiplas dimensões de um conceito e seus relacionamentos com outros conceitos, ampliando o entendimento acerca do conceito original.

3.2 Orientação na busca de informações

Decisional guidance se refere às funcionalidades de um sistema computacional interativo que têm o efeito de esclarecer, influenciar ou direcionar os usuários à medida que eles exercitam as escolhas que o sistema lhes permite (SILVER, 1991). Aplicado à busca de informações, a orientação pode incluir abordagens de navegação que ajudam os usuários a encontrar informações mais facilmente (LANKTON *et al.*, 2012). Nesta mesma linha, Lankton *et al.* (2012) sugerem que uma ferramenta de busca que permita a navegação participativa (busca por palavras-chave) aliada a uma orientação dinâmica (sugestões do sistema com base nas escolhas do usuário para que este escreva *queries* melhores) podem produzir melhores resultados de busca.

3.3 *Information Retrieval*

Information Retrieval (IR) é a área de pesquisa que trata da busca de documentos inteiros ou informações dentro de documentos. São utilizados por bibliotecas e universidades para prover acessos a livros e outras publicações, através de metadados e também por indexação dos textos completos. Nas ferramentas de busca da internet é utilizada a indexação das páginas visitadas pelos *crawlers* dos motores de busca.

3.4 Sistemas de recuperação de informação baseados em ontologias

Na busca baseada em ontologia, esta é utilizada para expandir a *query* original do usuário Neste caso, sinônimos ou palavras associadas por outras relações semânticas às palavras-chave originais, são adicionadas aos parâmetros de busca. A *query* expandida corresponde à interpretação do sistema em relação à real necessidade de informação do usuário com base no domínio representado pela ontologia. A *query* pode ser expandida, por exemplo, com descendentes e/ou ascendentes na hierarquia em relação aos conceitos originais, ou por instâncias destes níveis na ontologia.

Pesquisas têm medido os efeitos da expansão da *query* com base em ontologias (GULLA *et al.*, 2007; WASILEWSKI, 2011). Os efeitos são medidos em termos de melhora da precisão (mede a porcentagem dos documentos trazidos que são pertinentes) e cobertura (mede a porcentagem dos documentos pertinentes que são trazidos). Essas pesquisas indicam ainda que a expansão automáticas da *query* gera ganhos em precisão e cobertura no caso de *queries* originais curtas e pouco específicas ou vagas (em torno de duas ou três palavras), gerando pouco ganho quando a *query* original é mais completa e o usuário já expressou sua necessidade de informação de forma mais acurada, tal que a adição de termos relacionados não contribua muito. Os autores mencionam que, em geral, as *queries* dos usuários são curtas e que estes consideram a economia de expressões como mais importante que a possibilidade de especificar detalhadamente a necessidade de informação, visto que poucos utilizam as funcionalidades de busca avançada dos motores de busca. Neste caso, a estratégia de uso de estruturas ontológicas na reformulação das pesquisas se torna importante.

Alguns sistemas de busca enriquecidos por ontologia (GULLA *et al.*, 2007; CALEGARI; PASI, 2008; MOTTA *et al.*, 2000) apresentam uma interface que permite ao usuário selecionar os conceitos-chave na ontologia, construindo a *query* através da navegação na ontologia (seguindo as relações entre os conceitos) – esta navegação leva à criação de uma lista de parâmetros de busca que são ligados por operadores de agregação (“AND” e “OR”). Desta forma, a ontologia apoia o usuário na identificação dos conceitos a serem pesquisados, diferentemente do processo automático de expansão de *query* no qual o usuário tem um papel passivo no processo de expansão. O usuário pode navegar tanto no nível dos conceitos como no das instâncias da ontologia e, ao se selecionar um conceito, o mesmo é adicionado à *query* e o nível das instâncias é então envolvido no processo. Os operadores de agregação são também selecionados pelo usuário. Os autores defendem vantagens desse processo referentes à maior transparência no uso da ontologia.

No contexto da inteligência competitiva, a ontologia deve fornecer vocabulário referente às necessidades de monitoração (CAO, 2006), auxiliando na definição dos temas monitorados.

4 PROJETO

O projeto utiliza os conhecimentos da área de IR (*Information Retrieval*) na aplicação de ontologias para expansão semântica da busca de informações, aliados a conceitos de busca facetada, normalmente utilizada em bases de dados estruturadas, que explicita para o usuário as possíveis dimensões ou pontos de vista da informação procurada. O sistema facilita a aplicação de filtros das informações antes de se submeter a busca a um motor de busca. Para cada palavra de busca digitada, a ferramenta apresenta ao usuário sugestões de termos que podem ser adicionados para estreitar o escopo da busca de uma das formas a seguir:

a) Adicionando um conceito mais específico ao conceito original representado pela palavra digitada – uma espécie de *drill-down* de uma ferramenta OLAP;
b) Adicionando um conceito da mesma dimensão de análise com o qual o conceito original esteja relacionado na ontologia através de relação não hierárquica;
c) Adicionando um conceito de outra dimensão ou faceta do modelo com o qual o conceito original esteja relacionado na ontologia através de relação não hierárquica.

Quando a especificação de uma busca não é detalhada, um buscador comum age como se realizasse uma união das possíveis interpretações dos critérios de busca, ocasionando a sobrecarga de resultados. Quando o usuário digitar Oracle, por exemplo, pode estar se referindo à empresa fornecedora de *software* ou ao *software* de banco de dados – o significado não pode ser “desambiguado” sem a participação do usuário. A lógica de expansão do sistema deste trabalho então sugere as possibilidades de significado para que o usuário faça a escolha – neste caso pode-se expandir a busca para “banco de dados Oracle” ou “fornecedor Oracle”,

por exemplo. A funcionalidade é semelhante a um OLAP hierárquico apoiado em palavras de busca e uma estrutura provida pela ontologia.
A base de documentos volátil não permite a classificação ou mapeamento a priori de cada documento em função da ontologia – isto seria possível se o buscador fosse preparado para indexar os documentos em função da ontologia. Em vez disto, o sistema se vale da adição de palavras-chave originadas da ontologia para aumentar as chances de se obterem resultados relevantes na busca. Ainda que não garanta a aderência do conteúdo dos documentos voláteis ao vocabulário da ontologia, oferecer ao usuário uma maneira simples de melhor caracterizar sua pesquisa, no domínio da busca, tende a mitigar o problema da sobrecarga de informações.

4.1 Arquitetura da solução

O sistema é formado pelos três componentes descritos abaixo:

I. Tela de interface para busca – browser com página do Google ou outro motor de busca comum, onde o usuário percorre os passos listados abaixo, em sequência:
   - Digita as palavras para busca;
   - A cada palavra digitada, recebe do sistema uma lista de palavras adicionais sugeridas para expansão da *query*;
   - Escolhe na lista de expansão de *query* as palavras que melhor definem o contexto da busca original;
   - Interage para eventualmente fazer alterações manuais na expansão de busca feita automaticamente;
   - Submete as palavras da busca expandida para execução.

II. Componente mediador:
   - Recebe as palavras da busca inicial do usuário;
   - Procura pelos conceitos que as representam na ontologia;
   - Realiza a expansão dos termos originais com os conceitos da ontologia relacionados a eles, adicionando-os aos termos originais com operador lógico implícito “AND” ;
   - Retorna a busca expandida para a tela de interface.

   O componente mediador foi implementado com o software livre TypingAid, com o recurso “autocompletar” no campo de digitação da *query*. Para cada palavra digitada neste campo, o software faz a busca da mesma num arquivo texto preparado com as frases de expansão de busca de cada conceito existente na ontologia. Se o usuário selecionar uma das frases para expansão da *query*, a palavra original é substituída pela frase que contém a mesma palavra e as palavras adicionais.

III. Banco de dados com a ontologia do domínio armazenada na forma de triplas RDF (<sujeito> <predicado> <objeto>) exportado como arquivo texto com as possíveis expansões de buscas referentes a cada conceito da ontologia, para integração com o componente mediador.

As ilustração a seguir, mostram a aparência da interface da ferramenta utilizando o Google e o Regain (ferramenta *desktop search* baseada no Lucene), respectivamente:

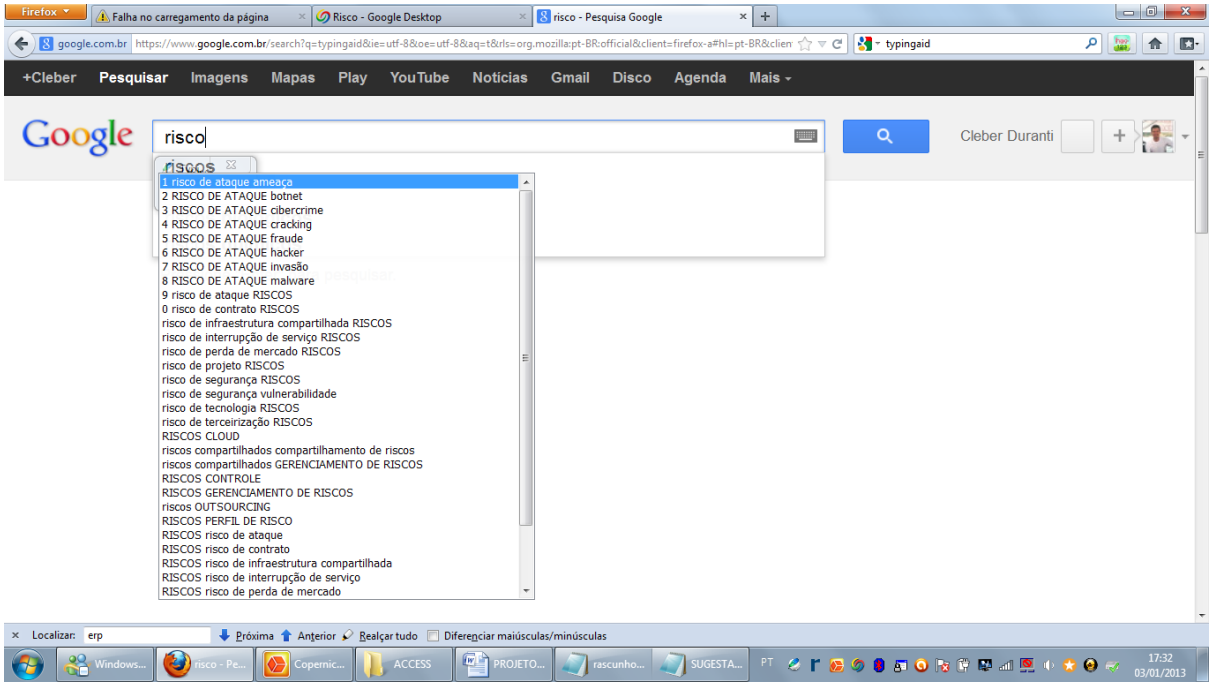

Ilustração 1 – Busca com o Google.

4.2 Ontologia "*outsourcing* de TI"

O CMap representa graficamente conceitos e relações e exporta o modelo para triplas RDF (<sujeito> <predicado> <objeto>) para armazenamento em banco de dados relacional.

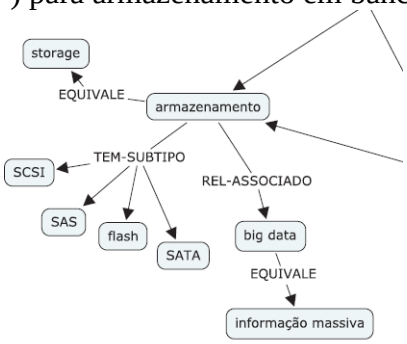

Ilustração 2 – Fragmento da ontologia "*outsourcing* de TI".

Na Ilustração 2, têm-se exemplos de relações de especialização/generalização ("flash" é um subtipo ou especialização de "armazenamento") e de uma relação de associação ("armazenamento" está associado ao conceito de "big data").

No topo da ontologia tem-se o conceito "*outsourcing* de TI" e, um nível abaixo na hierarquia, estão os grandes conceitos listados abaixo, referidos neste trabalho como dimensões de análise do "*outsourcing* de TI":

a) Tecnologia;
b) *Datacenter;*
c) Provedores (empresas que proveem os serviços de *outsourcing* de TI aos clientes);
d) Fornecedores (fornecedores dos provedores de *outsourcing* de TI);
e) Cliente (cliente do *outsourcing* de TI);
f) Recursos humanos;
g) Governança;
h) Motivadores (o que leva o cliente a terceirizar a TI);

i) Riscos;
j) Serviços (leque de serviços do *outsourcing* de TI);
k) Operação;
l) Recursos tecnológicos (subdividido em *software*, *hardware* e telecomunicações).

Essas dimensões de análise foram escolhidas em função da sua importância na monitoração do ambiente competitivo, como explicado adiante (ver item 4.2.1.1).

4.2.1 Construção da ontologia

Seguiu-se o tutorial para criação de ontologias da universidade de Stanford (NOY; MCGUINNESS, 2001), utilizando-se a abordagem mista para construção das hierarquias de classes: *top-down* e *botton-up* (Uschold e Gruninger,1996). A primeira ramificação do conceito topo da ontologia foi feita da forma *top-down* na definição das dimensões de análise do domínio, refletindo conceitos do modelo de Cadeia de Valor (ver item 4.2.1.1). Na direção *botton-up*, foram colhidos termos da base de notícias, de forma que o vocabulário da ontologia não ficasse descasado do vocabulário comum utilizado nas notícias da área (ver item 4.2.1.2). Os termos colhidos foram complementados e agrupados dentro das dimensões de análise, e as hierarquias criadas com apoio da literatura de outsourcing de TI e dos especialistas participantes da construção da ontologia.

4.2.1.1 Construção da ontologia "top-down"

Dos modelos de Cadeia de Valor (PORTER, 1985), Sistema de Valor (PORTER; 2008) e Cinco Forças Competitivas (PORTER, 1979), foram derivados os conceitos do segundo nível da ontologia (logo abaixo do conceito topo "*outsourcing* de TI"), aqui denominados "dimensões de análise", conforme descrito a seguir:

Modelo de Cadeia de Valor:

- A Infraestrutura foi representada pela dimensão **Datacenters**;
- A Gerência de Recursos Humanos foi representada pela dimensão **Recursos Humanos**;
- O Desenvolvimento de Tecnologia foi representado pela dimensão **Recursos Tecnológicos**;
- As Operações foram representadas pela dimensão **Operação**;
- Marketing e Vendas foram representados pela dimensão **Motivadores** do *outsourcing;*
- Serviços foram representados pela dimensão **Serviços**. Esta dimensão mapeia,o leque de serviços dos provedores de *outsourcing*;

Modelo de Sistema de Valor:

- Fornecedores foram representados pela dimensão **Fornecedores** na ontologia;
- Fabricante foi representado pela dimensão **Provedor** de outsourcing na ontologia;
- Varejista foi representado pela dimensão **Clientes** na ontologia -o elo final da cadeia.

Modelo das Cinco Forças Competitivas:

O modelo das Cinco Forças não acrescentou novas dimensões à ontologia, mas foi levado em consideração na criação dos conceitos pertencentes às dimensões. As seguintes forças foram consideradas: Fornecedores, Potenciais entrantes,. Compradores, Substitutos.
Verificou-se também que fossem mapeados na ontologia os atores típicos nos processos de monitoração do ambiente competitivo, conforme Lesca (2003): Concorrentes, Clientes, Fornecedores, Investidores, Parceiros, Poderes públicos, Grupos de pressão.

4.2.1.2 Construção da ontologia "botton-up"

Uma amostra de aproximadamente 5% das notícias sobre *outsourcing* de TI, disponibilizadas pelos principais sites nacionais de notícias de TI entre setembro de 2011 e agosto de 2012, foi analisada para extração manual de termos para a ontologia..

4.3 Banco de dados de trabalho do sistema

As triplas RDF (<sujeito> <predicado> <objeto>) da ontologia são armazenadas numa única tabela de banco de dados (tabela ONTOLOGIA), conforme o modelo de tabela vertical para representação e manipulação de ontologias (DEHAINSALA; PIERRA; BELLATRECHE, 2007). As demais tabelas são populadas a partir da tabela ONTOLOGIA e são utilizadas pelo módulo de expansão das buscas, armazenando os conceitos da ontologia (tabela CONCEITO), e os conceitos relacionados a eles nas deamis tabelas (tabelas SUPERTIPO, SUBTIPO, TODO, PARTE, EQUIVALENTE, RELACIONADO). A partir das tabelas CONCEITO, SUPERTIPO, SUBTIPO, TODO, PARTE, EQUIVALENTE e RELACIONADO, é populada a tabela SUGESTAO, a qual passa a já conter o *string* de expansão montado para cada conceito da ontologia.

4.4 Construção da interface

A interface para uso do sistema de expansão de busca com base em ontologia de domínio foi construída através da integração da base de dados em MS Access, que contém a ontologia em formato RDF e tabelas auxiliares, com o *software* Typingaid com recursos de *auto-complete* – um *software* livre responsável por exibir uma lista de expressões para substituição ou complementação de cada palavra digitada num campo de entrada de dados. Neste trabalho, esta funcionalidade é utilizada ao se digitar cada palavra-chave no campo de busca de um buscador padrão como o Google, por exemplo.

O conceito de *auto-complete* envolve que a interface preveja que palavras ou frases que o usuário deseja digitar. No caso deste trabalho, a previsão é feita no nível semântico, já que o sistema passa a prever o conjunto de palavras-chave que melhor definem a necessidade de informação do usuário, com base nas relações entre os conceitos da ontologia.

Para cada conceito existente na ontologia, as possíveis expansões são compostas, preliminarmente, através da adição dos conceitos relacionados ao conceito original. A conexão dos termos adicionados aos termos originais é feita através do operador lógico "AND" implícito.

Após a execução do *script* de criação das expansões, a tabela SUGESTAO passa a conter, para cada conceito, todas as possíveis expressões expandidas. Esta tabela é exportada para um arquivo texto que é então utilizado como base de sugestões de *auto-complete* do software Typingaid.

Na grafia dos termos sugeridos para expansão, são utilizadas letras maiúsculas ou minúsculas, dependendo da relação existente entre o termo original, digitado pelo usuário, e o termo sugerido para expansão, como descrito no quadro a seguir. Esta distinção visa a deixar explícito, para o usuário, se ele está transitando do conceito mais específico para o mais geral (subindo na hierarquia, uma espécie de *drill-up*) ou transitando do conceito mais geral para o mais específico (descendo na hierarquia, uma espécie de *drill-down*), ou ainda transitando para conceitos associados ao conceito original sem uma relação de hierarquia (relação "lateral" – uma espécie de *drill-across*).

4.5 Funcionamento do sistema

O funcionamento é similar ao do Google Suggest – sugestões de palavras que o Google oferece quando digitamos uma palavra de busca. O Google, porém, sugere palavras com base nas buscas mais populares, enquanto que neste protótipo as sugestões são feitas com base nos conceitos ontologia.relacionados às palavras digitadas.

Para cada termo digitado pelo usuário, o sistema de apoio procura por conceitos diretamente relacionados ao termo na ontologia (distância "1" na rede de conceitos que representa a ontologia). O sistema mostra então, ao usuário, uma ou mais *strings* compostas pela concatenação do conceito original com um conceito relacionado, separados por "espaço" (o espaço corresponde a um operador lógico "AND" implícito, na configuração original dos buscadores comuns). Dessa forma, o usuário vai sendo guiado a contextualizar melhor cada termo de busca, de forma a, ao submeter a pesquisa para execução no final deste processo, obter um conjunto de respostas menos extenso e com maior probabilidade de conter elementos relevantes. O exemplo a seguir ilustra o funcionamento do sistema:

Exemplo: Caso a palavra "Oracle" seja digitada na pesquisa original, será expandida da forma indicada no Quadro 1, em função das relações extraídas da ontologia, conduzindo o usuário a uma desambiguação de termos:

Quadro 1 – Exemplo: expansões do conceito "Oracle"

***Query* original**	***Query* expandida**	**Observação**
Oracle	• *ORACLE BANCO DE DADOS* • *ORACLE ERP* • *Oracle FORNECEDORES* • *ORACLE OPEN OFFICE* • *ORACLE Oracle10* • *ORACLE Oracle9* • *ORACLE Sun*	• Oracle como *software* de banco de dados • Oracle como *software* ERP • Oracle como fornecedor • Oracle como *software* Open Office (da Sun) • Ssubtipo de banco de dados Oracle • Ssubtipo de banco de dados Oracle • Sun como parte da empresa Oracle

4.6 Testes do sistema

O sistema foi testado de duas formas:

a) Realização de buscas simuladas pelo autor da pesquisa, sobre uma base de notícias baixadas da internet: descrita no item 4.6.1 adiante;

b) Realização de buscas na internet por um grupo de usuários que foram solicitados a utilizar a ferramenta e a responder ao questionário de avaliação: item 4.6.2.

4.6.1 Simulações de buscas

Para simulação das buscas com a ferramenta, foi formada uma base de notícias coletadas dos sites de notícias *online* acerca do mercado de TI no Brasil. A base foi populada pela extração manual do arquivo PDF das notícias disponibilizadas dentro do período de um ano (setembro de 2011 a agosto de 2012) dos principais sites de notícias *online*. Os arquivos PDF foram transformados em arquivos texto e as simulações foram então realizadas com estes arquivos no sistema de arquivos local do computador, rodando Windows e utilizando a ferramenta de busca Regain em substituição ao Google Desktop que não é mais suportado pela Google. O Google Desktop é similar ao buscador da Google para internet, porém, realizam as buscas nos arquivos locais do *desktop* em vez de realizar a busca na internet. O Regain tem a mesma funcionalidade, porém se utiliza do buscador de código aberto Lucene (KLEINER, 2011) e foi utilizado para as simulações.

As simulações foram realizadas sobre base de textos locais para permitir repetições nos testes sem problemas com a volatilidade da internet. Abaixo, o procedimento utilizado para preparação da base local de notícias.

- Busca por "outsourcing de TI" no Google em agosto de 2012, resultando nos seguintes números de conteúdos encontrados:
    - Qualquer data: 7.140.000
        - Último ano: 321.000

A partir desses dados, foram analisados os sites de noticiais usuais do mercado de TI e verificadas as quantidades de anúncios contendo o tema "*outsourcing* de TI" de cada fonte, conforme a tabela a seguir.

Tabela 1 – Fontes e volumes de notícias para as simulações

Site	Quantidade de notícias com "*outsourcing* de TI" no conteúdo com data entre setembro de 2011 e agosto de 2012
www.convergenciadigital.uol.com.br	231
www.baguete.com.br	148
www.info.abril.com.br	91
www.informationweek.itweb.com.br	81
www.computerworld.uol.com.br	70
www.cio.uol.com.br	51
www.metaanalise.com.br	31
www.tiinside.com.br	20
www.idgnow.uol.com.br	8
TOTAL	**731**

A seguir são mostrados os resultados de três simulações de buscas realizadas nesta base de notícias com o buscador Regain/Lucene em conjunto com a ferramenta de apoio à busca com base na ontologia "*outsourcing* de TI" desenvolvida neste trabalho.
A análise dos resultados das simulações mostra uma intersecção relativamente baixa dos resultados à medida que eram utilizados os termos sugeridos pela ferramenta para expansão da busca original, o que indica uma boa sensitividade dos termos da ontologia que, por sua vez, aumenta a precisão da *query* (KEKÄLÄINEN et al., 1998; NECIB; FREYTAG, 2005) .

4.6.1.1 Simulação – Busca pelo termo "Oracle" e expansões

A busca pelo termo "Oracle" resulta em 39 dos 731 arquivos da base de notícias, conforme mostrado no quadro a seguir:
Com a expansão da busca pela adição de termos sugeridos pela ferramenta, foram obtidos os resultados mostrados no quadro adiante (cada bloco mostra os resultados da busca com uma das possíveis expansões do conceito original "Oracle"):
O quadro a seguir faz uma representação da sensitividade dos termos da ontologia utilizados nas expansões da simulação de busca. A primeira coluna contém os títulos de todas as notícias trazidas com a busca original do conceito "Oracle" e as demais colunas indicam, nas interseções linha x coluna, quais notícias são trazidas para cada opção de expansão da busca sugerida (cada expansão é indicada pelos termos escritos na vertical):

Quadro 2 – Resultados das buscas expandidas do conceito "Oracle"

Query expandida	Númeor de resultados
Oracle ERP	3
Oracle banco de dados	5
Oracle Sun	4
Oracle servidor	8
Oracle fornecedor	11

Oi feita também uma análise de sensitividade das buscas, ou seja, uma análise de quão efetiva era a utilização dos termos adicionados à busca quanto à separação dos conjuntos de resultados resultados (suprimido do artigo por questão de espaço).

4.6.2 Levantamento da aceitação da tecnologia

Os seguintes temas foram sugeridos, como necessidades hipotéticas de informações, aos usuários para a realização dos testes de aceitação de tecnologia da ferramenta de busca para posterior resposta ao questionário adaptado do TAM3: Riscos do *outsourcing*; Oracle no mercado de *outsourcing*; Projetos em *Cloud*; Profissionais especialistas em *outsourcing*; Serviços disponíveis para *outsourcing*; Provedores de *outsourcing*; Tecnologias utilizadas no *outsourcing*.

Uma solicitação para execução do teste do sistema e resposta ao questionário foi enviada por *e-mail* para uma amostra por conveniência de **297** pessoas – profissionais e pesquisadores da área de TI. Os destinatários foram elencados entre contatos de grupo de estudos de Sistemas de Informação da Universidade e contatos profissionais do autor da pesquisa que atuam em diversas empresas da área de TI (*outsourcing*, gestão de projetos, desenvolvimento de software, departamento de TI de bancos, etc.). Desta forma, os resultados obtidos não podem generalizados.

Os testes dos usuários foram feitos ao longo dos meses de fevereiro e março de 2013 e foram obtidas, no total, **85** respostas ao questionário.

A tabela abaixo apresenta a média das avaliações para cada item. Nota-se a predominância de notas de avaliação melhores do que neutra (4: "neutro"), o que indica boa aceitação do sistema em todos os aspectos abordados no modelo.

Tabela 2 – Média dos graus de concordância com as afirmações do modelo

Grupo	Afirmação	Média das respostas
Utilidade percebida	1. O sistema melhora o desempenho nas buscas (facilita a formulação das buscas, lembrando facetas da informação que podem ser úteis)	5,69
	2. O sistema melhora a produtividade nas buscas (aumenta a proporção de resultados relevantes pelo aumento da seletividade da busca ao incentivar o uso de palavras mais específicas)	5,72
	3. O sistema melhora a efetividade das buscas (buscas mais assertivas pela contextualização provida pela ferramenta)	5,54
	4. Considero o sistema útil	5,69
Facilidade de uso percebida	5. A interação com o sistema é clara e compreensível	5,33
	6. Interagir com o sistema não requer muito esforço mental	5,82
	7. É fácil usar o sistema	5,82
	8. É fácil fazer o sistema executar o que quero (facilidade na operação do sistema)	5,48
Autossuficiência	9. Eu poderia utilizar o software se não houvesse ninguém por perto para me dizer como fazer	4,88
	10. Eu poderia utilizar o software se alguém me mostrasse antes como fazer	5,08
	11. Eu poderia utilizar o software se eu tivesse usado pacotes similares antes	4,88
Percepção de controle	12. Eu tenho controle sobre o sistema	5,24
	13. Eu tenho os recursos necessários para usar o sistema (Equipamento, software)	6,33
	14. Dados os recursos, oportunidades e conhecimento para utilizar o sistema, seria fácil utilizá-lo	6,24
	15. O sistema é compatível com outros sistemas que eu uso	5,66
Percepção de prazer	16. Considero o sistema agradável de usar	5,34
Qualidade do resultado	17. A qualidade do resultado que obtenho do sistema é alta	5,33
	18. Não tenho problema com a qualidade do resultado do sistema	5,42
	19. Considero os resultados do sistema excelentes	5,20

Demonstrabilidade do resultado	20. Não teria dificuldade para contar aos outros sobre os resultados do uso do sistema	5,91
	21. Creio que eu poderia comunicar aos outros as consequências do uso do sistema	5,84
	22. Os resultados do uso do sistema são aparentes para mim	5,80
	23. Não teria dificuldade em explicar por que usar o sistema pode trazer ou não benefícios	5,86
Intenção comportamental	24. Você estaria disposto a utilizar o sistema se tivesse acesso a ele e necessitasse se manter a par das notícias de outsourcing de TI (situação hipotética)	5,78
	MÉDIA geral	**5,58**

4.7 Discussão

A avaliação do protótipo feita pelos usuários indicou que a ferramenta de expansão interativa de busca, com base na ontologia do domínio de negócio-alvo, ajuda na obtenção de informações mais seletivas nas buscas da internet.

A ontologia de domínio de negócio construída manualmente, incorporando modelos de competitividade, de acordo com as avaliações, mostrou-se útil como subsídio para seleção de notícias, ainda que estas façam parte de uma base dinâmica que dificulta o perfeito casamento entre os termos da ontologia e os termos das notícias. A construção da ontologia a partir do conhecimento do negócio e utilizando uma amostra de notícias para alinhamento de vocabulário se mostrou eficiente.

Apesar de os resultados não poderem ser generalizados, o sistema proposto apresenta-se como uma ferramenta útil para mitigação da sobrecarga de informações na internet, ao procurar estruturar informações de natureza não estruturada, contribuindo para um maior controle sobre o que é resgatado das bases *online* de notícias, por conduzir os usuários a especificarem melhor as suas pesquisas.

5 REFERÊNCIAS BIBLIOGRÁFICAS

BRENNAN, L. L. The Scientific Management of Information Overload. **Journal of Business and Management**, n. 1997, p. 121-135, 2006.

CALEGARI, S.; PASI, G. **Personalized Ontology-Based Query Expansion**2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. **Anais**. Ieee, dez. 2008. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4740774>. Acesso em: 25 jul. 2012

CAO, T. D. **Exploitation du web sémantique pour la veille technologique**. [S.l.] UNIVERSITE de Nice-Sophia Antipolis, 2006.

CHUNG, W.; CHEN, H.; JR, J. A. Y. F. N. A Visual Framework for Knowledge Discovery on the Web : An Empirical Study of Business Intelligence Exploration. **Journal of Management Information Systems**, v. 21, n. 4, p. 57-84, 2005.

COHEN, W. M.; LEVINTHAL, D. A. Absorptive Capacity : A New Perspective on Learning and Innovation. **Administrative Science Quarterly**, v. 35, n. 1, p. 128-152, 1990.

CROSSAN, M. M. et al. An Organizational Learning Framework: From Intuition to Institution. **The Academy of Management Review**, v. 24, n. 3, p. 522-537, 1999.

DAFT, R. L.; WEICK, K. E. Toward a model of organizations as Interpretation systems. **The Academy of Management Review**, 1986.

DAVIS, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. **MIS Quarterly**, v. 13, n. 3, p. 319-340, 1989.

DAVIS, N. Information Overload , Reloaded contemporary information architecture can help reframe how we approach this thing called information overload . **Bulletin of the American Society for Information Science and Technology**, p. 45-50, 2011.

DEHAINSALA, H.; PIERRA, G.; BELLATRECHE, L. **OntoDB : An Ontology-Based Database for Data Intensive Applications.** Proc . of Database Systems for Advanced Applications (DASFAA ' 2007). **Anais**...2007

DENTON, D. K.; RICHARDSON, P. Using Intranets to Reduce Information Overload. **Journal of Strategic Innovation and Sustainability**, v. 7, n. 3, p. 84-95, 2012.

EPPLER, M. J.; MENGIS, J. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. **The Information Society**, v. 20, n. 5, p. 325-344, nov. 2004.

GARSHOL, L. M. Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. **Journal of Information Science**, v. 30, n. 4, p. 378-391, 1 ago. 2004.

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

GULLA, J. A.; BORCH, H. O.; INGVALDSEN, J. E. **Ontology Learning for Search Applications.** OTM'07 Proceedings of the 2007 OTM Confederated international conference on On the move to meaningful internet systems: CoopIS, DOA, ODBASE, GADA, and I. **Anais**...2007

HWANG, W.; SALVENDY, G. Effects of an ontology display with history representation on organizational memory information systems. **Ergonomics**, v. 48, n. 7, p. 838-58, 10 jun. 2005.

JENKIN, T. A. **Using information technology to support the discovery of novel knowledge in organizations**. [S.l.] Queen's University, 2008.

JONKER, C. M.; RIEMSDIJK, M. B. VAN; VERMEULEN, B. **Shared Mental Models - A Conceptual Analysis.** Proceedings of the 9th International Workshop on Coordination, Organization, Institutions and Norms in Multi-Agent Systems. **Anais**...2010

KEKÄLÄINEN, J. et al. **The impact of query structure and query expansion on retrieval performance.** Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval. **Anais**...1998

KLEINER, E. **Google Desktop (Search) & Regain**Seminar talk: Information Retrieval. **Anais**...Konstanz, Germany: University of Konstanz, 2011. Disponível em: <http://www.inf.uni-konstanz.de/fileadmin/dbis/ir_w10/Eike_Kleiner_GDS_and_Regain_Introduction.pdf>

KUDRYAVTSEV, D. **Mapping ontologies and contexts : from theory to a case study .**C&O-2006. **Anais**...Riva del Garda, Italy: 2006

KUHLTHAU, C. C. Inside the search process: Information seeking from the user's perspective. **Journal of the American Society for Information Science**, v. 42, n. 5, p. 361-371, jun. 1991.

LANKTON, N. K.; SPEIER, C.; WILSON, E. V. Internet-based knowledge acquisition: Task complexity and performance. **Decision Support Systems**, v. 53, n. 1, p. 55-65, abr. 2012.

LI, T. An Investigation and Analysis of Information Overload in Manager's Work. **iBusiness**, v. 03, n. 01, p. 49-52, 2011.

LI, Y. et al. A two-stage decision model for information filtering. **Decision Support Systems**, v. 52, n. 3, p. 706-716, fev. 2012.

MAES, P. Agents that reduce work and information overload. **Communications of the ACM**, v. 37, n. 7, p. 30-40, 1 jul. 1994.

MARSHALL, B. et al. EBizPort: Collecting and analyzing business intelligence information. **Journal of the American Society for Information Science and Technology**, v. 55, n. 10, p. 873-891, ago. 2004.

MOTTA, E.; BUCKINGHAM SHUM, S.; DOMINGUE, J. Ontology-driven document enrichment: principles, tools and applications. **International Journal of Human-Computer Studies**, v. 52, n. 6, p. 1071-1109, 2000.

NECIB, C. B.; FREYTAG, J. Semantic Query Transformation Using Ontologies. **9th International Database Engineering & Application Symposium (IDEAS'05)**, p. 187-199, 2005.

NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101 : A Guide to Creating Your First Ontology** (S. A. McIlraith, D. Plexousakis, & F. Harmelen, Eds.)**Development**. [S.l.] Citeseer, 2001. Disponível em: <http://bmir.stanford.edu/file_asset/index.php/108/BMIR-2001-0880.pdf>.

ONG, T.-H. et al. Newsmap: a knowledge map for online news. **Decision Support Systems**, v. 39, n. 4, p. 583-597, jun. 2005.

PORTER, M. E. How competitive forces shape strategy. **Harvard Business Review**, v. 57, n. 2, p. 137-146, 1979.

PORTER, M. E. **Competitive Advantage: Creating and Sustaining Superior Performance**. [S.l.] Free Press, 1985. v. 1 st edn

PORTER, M. E. **On Competition, Updated and Expanded Edition**. [S.l.] Harvard Business School Press, 2008.

SILVER, M. S. Decisional Guidance for Computer-Based Decision Support. **MIS Quarterly**, v. 15, n. 1, p. 105-122, 1991.

STUDER, R.; BENJAMINS, V. R.; FENSEL, D. I DATA & KNOWLEDGE Knowledge Engineering : Principles and methods. **Data & Knowledge Engineering - Special jubilee issue: DKE 25**, v. 25, n. 1-2, p. 161-197, 1998.

TATE, C. C. **Using Visualization Tools to Mitigate Information Overload on the Internet**. [S.l.] Georgetown University, 2008.

USCHOLD, M.; GRUNINGER, M. Ontologies: Principles, methods and applications. **Knowledge Engineering Review**, v. 11, n. 2, p. 93-136, 1996.

VAISHNAVI, V.; KUECHLER, B. Design Science Research in Information Systems Overview of Design Science Research. **Association for Inormation Systems**, n. 1978, p. 1-16, 2004.

VENKATESH, V.; BALA, H. Technology Acceptance Model 3 and a Research Agenda on Interventions. **Decision Sciences**, v. 39, n. 2, p. 273-315, 2008.

VILLAGE, D. Cyberspace 2000 : Dealing with Information Overload. **COMMUNICATIONS OF THE ACM**, v. 40, n. 2, p. 19-24, 2000.

WASILEWSKI, P. **Query Expansion by Semantic Modeling of Information Needs.** Proceedings of the international workshop CS&P 2011. **Anais**...2011. Disponível em: <http://csp2011.mimuw.edu.pl/proceedings/PDF/CSP2011523.pdf>

WU, H. et al. Mining web navigations for intelligence. **Decision Support Systems**, v. 41, n. 3, p. 574-591, mar. 2006.

ZACK, M. H.; HALL, H. Developing a Knowledge Strategy. **California Management Review**, v. 41, n. 3, p. 1-18, 1999.