Global Journal of Computer Science and Technology: H
Information & Technology
Volume 15 Issue 6 Version 1.0 Year 2015
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Selection of Online News for Competitive Intelligence: use of Business Domain Ontology for Internet Search Semantic Query Expansion

By Cleber Marchetti Duranti & Fernando Carvalho De Almeida
University of São Paulo, Brazil

Abstract- The Internet provides ever increasing volumes of news and information about the environment in which companies operate. This can lead to information overload, in which the volume of information available overwhelms the processing power of the user. Methods and tools that help separate potentially useful information from irrelevant information need to be developed. This research applied design research to investigate the development of a tool to help users refine internet searches on competitive intelligence. It used modeling of the target business area in the form of anontology to aid the formulation of search terms through interactive semantic expansion of the keywords entered by users.

Keywords: *information retrieval, ontology, search engines, competitive intelligence, design research*

GJCST-H Classification: *I.2.11*

SELECTIONOFONLINENEWSFORCOMPETITIVEINTELLIGENCEUSEOFBUSINESSDOMAINONTOLOGYFORINTERNETSEARCHSEMANTICQUERYEXPANSION

Strictly as per the compliance and regulations of:

Selection of Online News for Competitive Intelligence: Use of Business Domain Ontology for Internet Search Semantic Query Expansion

Cleber Marchetti Duranti [α] & Fernando Carvalho De Almeida [σ]

Abstract- The Internet provides ever increasing volumes of news and information about the environment in which companies operate. This can lead to information overload, in which the volume of information available overwhelms the processing power of the user. Methods and tools that help separate potentially useful information from irrelevant information need to be developed. This research applied design research to investigate the development of a tool to help users refine internet searches on competitive intelligence. It used modeling of the target business area in the form of anontology to aid the formulation of search terms through interactive semantic expansion of the keywords entered by users.

Keywords: information retrieval, ontology, search engines, competitive intelligence, design research.

I. Introduction

The Internet represents a rich external resource of information about the environment and is used extensively by organizations (Marshall et al. 2004). Researchers have pointed out, however, the difficulty in locating relevant information from the vast amount available online. This is the problem of information overload (Chung et al. 2005), which is experienced, for example, when a user searches for information on a given topic through a search engine and gets a long list of results. It is a standing problem for companies that use the Internet as a key source of information (Davis 2011; Denton & Richardson 2012; Jenkin 2008; Li 2011; Li et al. 2012; Tate 2008).

In an extensive review of the literature on information overload, Eppler and Mengis (Eppler & Mengis 2004) suggest that overloading occurs when the information processing requirement exceeds the processing capacity of the individual or organization. Processing encompasses the collection, interpretation, and synthesis of information in the context of the organization's decision-making needs.

Information overload is a consequence of both the abundance of information and deficits in the applied filters. It can be addressed by the field of information architecture (Davis 2011). As more information becomes available, users require better tools to help them filter the flow of information and find items of interest (Maes 1994). There will be no final solution to information overload but rather cycles of refinement and improvement (Maes 1994).

Understanding and being updated on the external environment in which companies operate demands the discovery of knowledge through individual and organizational learning processes (Jenkin 2008). As individuals have a limited capacity to assimilate new information, they build meanings selectively by focusing on information that connects with that which they already know (Kuhlthau 1991). The learning of new concepts must be founded on familiar knowledge and mental models (Cohen & Levinthal 1990), which are the structures that help simplify and organize information (Crossan et al. 1999). They comprise structures that represent knowledge as a network of abstract concepts with attributes, values, relationships, and rules. Both individuals and organizations have mental models. In the case of an organization, the mental model is an understanding shared and negotiated by its members.

In information science, an ontology expresses the consensus knowledge of a domain. The concepts that fall within the area are represented as nodes in a network, and relationships between concepts are represented by arcs, which depict the type of relationship. An explicit specification of a conceptualization" (Gruber 1995) is the means of representing shared mental models (Jonker et al. 2010; Kudryavtsev 2006). An ontology describes the common knowledge of a group about a specific area in a format that can be processed by a machine and defines its concepts, properties, and attributes in a vocabulary common to the group. The ontology can play a crucial role in establishing both explicit individual mental models and shared mental models (Hwang & Salvendy 2005) within an organization. This explicit representation of the competitive environment in the form of an ontology can support the acquisition of new information about the environment and assist in incrementing or updating the organization's current view.

This study describes the construction of a system to support the search for and selection of information on the Internet by using an ontology representative of the company's business domain. This was based on the semantic expansion of search terms defined by the user when searching for online news using standard search engines such as Google. The

Author: University of São Paulo. e-mail: clebermd@uol.com.br

expansion was designed to add terms to the search words entered by the user and enhance the context of the search, thus improving the quality of the results. The system increased the chances of finding information relevant to the subject in focus and of avoiding information overload.

a) Research problem and purpose

The research aimed to explore the application of an ontology of a business domain in order to increase the selectivity of information searches related to the competitive business environment.

i. *Specific Objectives*

- O1 - Construction of a domain ontology pilot "IT outsourcing";
- O2 - Construction of a system to support internet searches by making use of the relationships between the concepts of the ontology for the semantic expansion of search words;
- O3–Evaluation of the system using the Technology Acceptance Model (TAM3).

ii. *Research Questions*

- Q1–Is a manually constructed business domain ontology incorporating competitive models useful as a resource for news selection (dynamic database)?

This question is addressed the specific objectives O1 and O3.

- Q2–Does the use of ontological relationships to expand the search terms increase the selectivity of the information retrieved?

This question is addressed by all three specific objectives O1 to O3.

- Q3 - How can the business domain ontology be used to reduce information overload?

This question is addressed by the specific objectives O2 and O3.

II. Literature Review

a) Information overload in the Internet

Information overload means that more information is available than can be acquired, processed, stored, or redeemed (Brennan 2006). It arises when the supply exceeds the capacity to consume (Eppler & Mengis 2004) and results from the possibility of capturing and accessing large volumes of data made available by information technology (Ong et al. 2005). The problem lies not in the abundance of information but in the failure to filter that information. The ease and low cost of publishing on the Internet have moved the quality filter downstream (Davis 2011). Search engines represent the first attempt to deal with information overload on the web but are currently seen as primitive (Village 2000).

Organizational learning theories can be applied to the construction of tools for knowledge discovery on the Internet (Jenkin 2008). Tools that incorporate the shared mental model of individuals in an organization can support incremental learning based on existing knowledge. These tools, in the form of ontologies and other semantic web technologies, can guide the acquisition of knowledge, particularly incremental acquisition, by supporting the exploration of multiple dimensions of a concept and its relationships with other concepts, thus enhancing understanding of the original (Jenkin 2008).

Absorptive capacity (Cohen & Levinthal 1990) concerns the ability of a firm to recognize the value of new external information, to assimilate it, and to make use of it for commercial purposes. However, this is a function of the previous stock of related knowledge. A crucial precondition for a company's capacity to innovate is the ability to exploit external knowledge but it is precisely the stock of existing knowledge that allows it to recognize the value of new information. The categories into which the preliminary knowledge is organized, the differentiation of categories, and the relationships between them are the tools that allow individuals to create meaning, and consequently, to acquire new knowledge (Bower & Hilgard 1981).

b) Guidance in seeking information

Decisional guidance refers to the features of an interactive computer system that clarify, influence, or direct users as they exercise choice (Silver 1991). Within information search, the guidance includes the navigational approaches that help users find information more easily (Lankton et al. 2012). Search tools that allow participatory navigation (search by keywords), combined with a dynamic orientation (suggestions offered by the system, based on previous user choices), can improve search results (Lankton et al. 2012).

c) Ontology-based information retrieval systems

In ontology-based searches, an ontology is used to expand the user's original query by exploiting semantic relations to add synonyms, or words associated with the original keywords, to the search parameters. The expanded query corresponds to the interpretation of the system, based on the user´s real information needs, within the domain represented by the ontology. The query may be expanded using descending and / or ascending concepts in the hierarchy, or instances of these levels in the ontology.

Researchers have investigated the effects of such ontology-based query expansions (Gulla et al. 2007), using measures such as improved accuracy (the percentage of all retrieved documents that are relevant) and coverage (the percentage of all relevant documents that are retrieved). These surveys suggest that automatic query expansion enhances accuracy and coverage when the original query was short (about two or three words), insufficiently specific, or vague but had little benefit when the original query was more complete and accurate. In such cases, the addition of related terms contributes little to the search. The authors report that user queries are often brief, as economy of expression is preferred to detailed specification of information needs as few users make use of the advanced search features of search engines. This makes the use of ontological structures in the reformulation of searches more important.

In the context of competitive intelligence, ontology should provide vocabularies related to monitoring needs (Cao 2006), thus assisting in the definition of the subjects to be monitored.

III. Research Methodology

This section presents the Design Research methodology used in this study, the methodology for building ontologies used in the construction of an "IT outsourcing" ontology, and the model for technology acceptance used to evaluate the prototype developed in the research.

a) Design Research

Design Research or Design Science Research addresses learning by building artifacts. The design itself (artifact construction) is used as a research method or technique (Vaishnavi & Kuechler 2004). It involves the design of new devices and the analysis of their use and / or performance to improve and understand the behavioral aspects of Information Systems.

This research applied the Design Science Research method to the construction of two artifacts: an ontology and a system for query expansion based on that ontology. These were proposed as countermeasures to information overload when searching for news on the Internet. Within the Design Research approach, a proposed solution is presented as being representative of a class of solutions for a class of problems.

b) Methodology for building ontologies

The methodology that was used for creating ontologies was taken from the Knowledge Systems Laboratory at Stanford University (Noy & Mcguinness 2001). It can be summarized as entailing the following steps.

i. *Determine the scope of the ontology by defining*
   - The area to be covered by the ontology;
   - The intended use of the ontology;
   - The type of questions that the information in the ontology should provide answers to;
   - The users and maintainers of the ontology.

ii. *Consider reusing existing ontologies from libraries of reusable ontologies such as*
   - Ontolíngua (http://www.ksl.stanford.edu/software/ontolingua/);
   - DAML (http://www.daml.org/ontologies/);
   - UNSPSC (www.unspsc.org);
   - RosettaNet (www.rosettanet.org);
   - DMOZ (www.dmoz.org).

However, the reuse of preexisting ontologies is challenging (Cao 2006) because consistency in conceptualization is required between the existing ontology and the desired one. Each ontology is dedicated to a specific purpose, and automatic import of vocabularies is impossible.

iii. *List the important terms in the ontology to create a preliminary list of concepts without worrying about the overlap and relationships between them, the properties that the concepts may have, or whether the concepts are classes or properties of classes.*

iv. *Define the classes and the hierarchy of classes. Several approaches are available (Uschold & Gruninger 1996), including.*
   - Top-down, wherein development begins with the definition of the most general concepts
   - Bottom-up, wherein development starts from the definition of the most specific classes or leaves of the hierarchy, before grouping these classes into more general concepts
   - A combination of top-down and bottom-up.

v. *Set the properties (slots) that describe the internal structure of concepts.*

vi. *Set the facets of the slots —data type, allowed values, cardinality, etc.*

vii. *Create instances of the classes—define the individuals represented by the classes by assigning values to the slots.*

c) *Technology Acceptance Model*

The TAM was developed to predict the adoption and use of new IT systems (Davis 1989). It proposes that the individual intention to use a technology is determined by two beliefs: perceived usefulness, i.e., extent to which a person believes that using a technology will enhance job performance and perceived ease of use, i.e., degree to which a person believes that the use of the technology will be effortless. TAM3, the most recent version of the model (Venkatesh & Bala 2008), has been adapted for the evaluation of the prototype in this research.

We were conducting a proof of concept rather than the introduction of a real software system into a work environment, therefore the TAM3 has been adapted for the evaluation of the prototype in this research. The Figure 1 shows how the TAM3 was adapted (see the "Adaptation if any" column) and the correspondence between the statements and the variables of this study.

We applied simulation tests to allow users to try the tool, using Likert-type scales in which users were asked to indicate on a scale of one to seven their agreement with each of the 24 items (V1-V24) as follows: 1: strongly disagree, 2: moderately disagree, 3: somewhat disagree, 4: neutral; 5: somewhat agree; 6: moderately agree; 7: strongly agree. The final questionnaire is given in

Table 1: (item 4.6)

Constructs	Statements from the original model	Adaptation if any	Variable
Perceived Usefulness (PU)	PU1 Using the system improves my performance in my job.	Unchanged	V1
	PU2 Using the system in my job increases my productivity.	Unchanged	V2
	PU3 Using the system enhances my effectiveness in my job.	Unchanged	V3
	PU4 I find the system to be useful in my job.	Unchanged	V4
Perceived Ease of Use (PEOU)	PEOU1 My interaction with the system is clear and understandable.	Unchanged	V5
	PEOU2 Interacting with the system does not require a lot of my mental effort.	Unchanged	V6
	PEOU3 I find the system to be easy to use.	Unchanged	V7
	PEOU4 I find it easy to get the system to do what I want it to do.	Unchanged	V8
Computer Self-Efficacy (CSE)	I could complete the job using a software package . . .		
	CSE1 . . . if there was no one around to tell me what to do as I go.	Unchanged	V9
	CSE2 . . . if I had just the built-in help facility for assistance.	Supressed as the pilot had no built-in help	
	CSE3 . . . if someone showed me how to do it first.	Unchanged	V10
	CSE4 . . . if I had used similar packages before this one to do the same job.	Unchanged	V11
Perceptions of External Control (PEC)	PEC1 I have control over using the system.	Unchanged	V12
	PEC2 I have the resources necessary to use the system.	Unchanged	V13
	PEC3 Given the resources, opportunities and knowledge it takes to use the system, it would be easy for me to use the system.	Unchanged	V14
	PEC4 The system is not compatible with other systems I use.	The word “not” was supressed for	V15

		clarity	
Computer Playfulness (CPLAY)	The following questions ask you how you would characterize yourself when you use computers:	Supressed as all testers work in the IT industry	
	CPLAY1 . . . spontaneous.		
	CPLAY2 . . . creative.		
	CPLAY3 . . . playful.		
	CPLAY4 . . . unoriginal.		
Computer Anxiety (CANX)	CANX1 Computers do not scare me at all.		
	CANX2 Working with a computer makes me nervous.		
	CANX3 Computers make me feel uncomfortable.		
	CANX4 Computers make me feel uneasy.		
Perceived Enjoyment (ENJ)	ENJ1 I find using the system to be enjoyable.	Unchanged	V16
	ENJ2 The actual process of using the system is pleasant.	Supressed	
	ENJ3 I have fun using the system.	Supressed	
Objective Usability	It was measured as a ratio of time spent by the subject to the time spent by an expert on the same set of tasks.	Supressed	
Subjective Norm (SN)	SN1 People who influence my behavior think that I should use the system.	Supressed	
	SN2 People who are important to me think that I should use the system.	Supressed	
	SN3 The senior management of this business has been helpful in the use of the system.	Supressed	
	SN4 In general, the organization has supported the use of the system.	Supressed	
Voluntariness (VOL)	VOL1 My use of the system is voluntary.	Supressed	
	VOL2 My supervisor does not require me to use the system.	Supressed	
	VOL3 Although it might be helpful, using the system is certainly not compulsory in my job.	Supressed	
Image (IMG)	IMG1 People in my organization who use the system have more prestige than those who do not.	Supressed	
	IMG2 People in my organization who use the system have a high profile.	Supressed	
	IMG3 Having the system is a status symbol in my organization.	Supressed	
Job Relevance (REL)	REL1 In my job, usage of the system is important.	Supressed	
	REL2 In my job, usage of the system is relevant.	Supressed	
	REL3 The use of the system is pertinent to my various job-related tasks.	Supressed	
Output Quality (OUT)	OUT1 The quality of the output I get from the system is high.	Unchanged	V17
	OUT2 I have no problem with the quality of the system's output.	Unchanged	V18
	OUT3 I rate the results from the system to be excellent.	Unchanged	V19
Result Demonstrabi-lity (RES)	RES1 I have no difficulty telling others about the results of using the system.	Unchanged	V20
	RES2 I believe I could communicate to others the consequences of using the system.	Unchanged	V21
	RES3 The results of using the system are apparent to me.	Unchanged	V22
	RES4 I would have difficulty explaining why using the system may or may not be beneficial.	Unchanged	V23
Behavioral Intention (BI)	BI1 Assuming I had access to the system, I intend to use it.	Unchanged	V24
	BI2 Given that I had access to the system, I predict that I would use it.	Supressed	
	BI3 I plan to use the system in the next <n> months.	Supressed	
Use (USE)	USE1 On average, how much time do you spend on the system each day?	Supressed	

Figure 1 : Constructs of the Technology Acceptance Model 3 and the variables of this research

i. *Survey validation*

To allow comparison between the factors in the TAM3 conceptual model and those observed in this study, a factorial analysis of the survey variables (corresponding to the twenty-four questions) was performed to verify the consistency of the results. The main objective of this study was not, however, to verify the dependency relationships between the constructs of TAM3.

IV. Project

The proposed solution used knowledge of the Information Retrieval (IR) area in applying ontologies for semantic expansion of information searches, combined with the faceted search that is widely used in structured databases. These make the possible dimensions or views of the requested information explicit to the user. The system facilitated the application of information filters before the submission of the search query. For each typed search term, the tool suggested additional terms to narrow the scope of the search in one of the following ways:

a) By adding a more specific concept to the original concept, which is equivalent to drill-down of an online analytical processing (OLAP) tool.

b) By adding a more general concept to the original concept, which is equivalent to drill-up of an OLAP tool.
c) By adding a concept of the same analysis dimension to which the original concept was related in the ontology, through some non-hierarchical relationship, equivalent to drill-across of a relational online analytical processing (ROLAP) tool.
d) By adding a concept from another dimension or facet of the model with which the original concept was related in the ontology, through some non-hierarchical relationship - also a sort of drill-across of a ROLAP tool.

When the specification of a search is not detailed, most browsers work as if performing a union of all the possible interpretations of the search criteria, leading to an overload of results. In the context of information technology, when the user types “Oracle,” for example, the intended reference may be to (a) the software provider company or to (b) the database software. The meaning cannot be “disambiguated” without user participation, and thus a standard search engine must consider a union of these possible meanings (a U b). The expansion logic used in this research stresses the significance of the user making the choice, in this example between expanding the search to “Oracle Database” or “Oracle provider”.

a) Architecture

The system comprised the three components described below and illustrated in Figure 2:

i. An interface window: this was a browser window with a Google page or other regular search engine that executed the following steps in the given sequence:
   - The user typed in the terms of the search;
   - For each term typed, a list of additional words for the expansion of the query was suggested;
   - From the query expansion list, the user chose the terms that better defined the context of the intended search;
   - Manual changes in the search expansion were made automatically;
   - The user submitted the expanded search terms.

ii. A mediator component which:
   - Received the words of the user's initial search;
   - Searched for concepts to represent them in the ontology;
   - Expanded the original terms with related concepts from the ontology;
   - Added these to the original terms with the implicit logical operator "AND";
   - Returned the expanded search terms to the interface.

The mediator component was implemented through an adaptation of the free software TypingAid (www.autohotkey.com), which enables autocomplete in the query typing field, using suggestions taken from a preloaded text file. When presented with a typed word, the software searches for the word inside the text file. In the prototype, the text file was preloaded with search expansion phrases, using the relationships between concepts in the ontology. If the user selected one of the phrases suggested for expansion of the query, the original word was replaced by a group of words containing the original word and the additional ones.

iii. A database with the domain ontology stored as Resource Description Framework (RDF) triples (<subject> <predicate> <object>) and exported as a text file containing the possible search expansions for each concept of the ontology for integration with the mediator component.

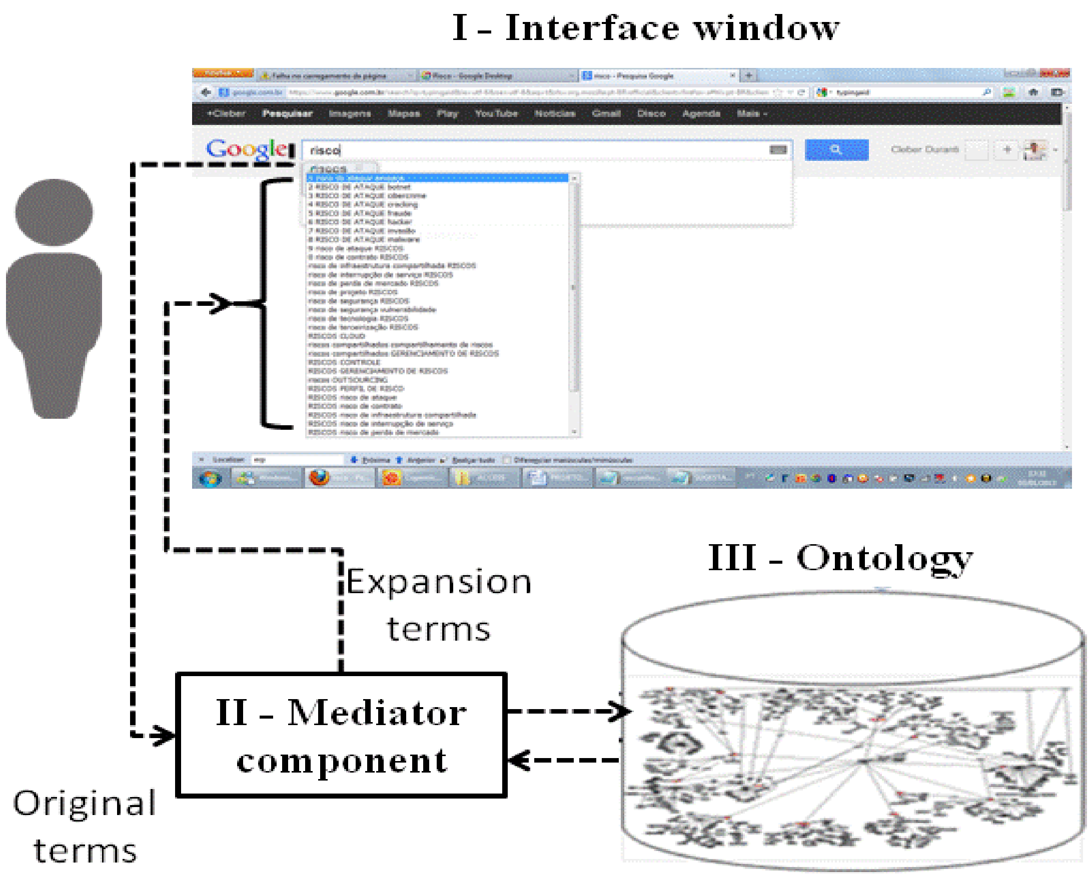

Figure 2 : Components of the solution

b) *Ontology "outsourcing"*

The ontology was designed using the Cmap software, which graphically represents concepts and relationships and exports the model as RDF triples (<subject> <predicate> <object>) to be stored in a relational database.

The Figure 3 gives examples of the relations of specialization / generalization ("flash" is a subtype or specialization of "storage") and association ("storage" is associated with the concept of "big data").

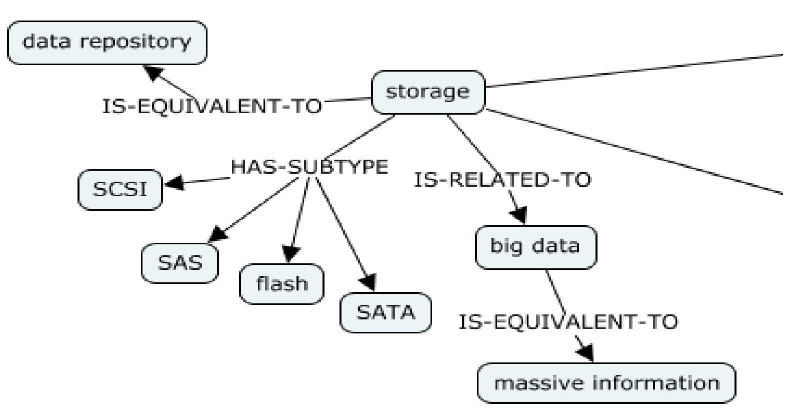

Figure 3 : Fragment of the ontology "outsourcing de TI"

As illustrated in Figure 4, the top level of the ontology contained the concept "IT outsourcing" and the second level contained the major concepts (referred to in this study as analysis dimensions).These were Technology, Datacenter, Providers (companies that provide IT outsourcing services to customers), Suppliers (suppliers to IT outsourcing providers), Clients (IT outsourcing customers), Human Resources, Governance, Drivers (which lead the customer to outsource IT), Risks, Services (range of IT outsourcing services), Operation, and Technological Resources (subdivided into software, hardware, and telecommunications).

These analytical dimensions were chosen based on their importance in monitoring the competitive environment, as explained in Section 4.2.1.1.

i. *Construction of the ontology*

The ontology was designed by the authors of the study and two other experts in the field, following the tutorial for creating ontologies from Stanford University (Noy & Mcguinness 2001) and using a mixed approach (top-down and bottom-up) for the construction of the class hierarchies (Uschold & Gruninger 1996). The constructed ontology was a light-weight one- an ontology for search engines on the Internet that consists of hierarchies of topics, giving less consideration to the strict definitions of the concepts and their organization (Mizoguchi 2003). This is adequate for applications in search expansion where the side relations (non-hierarchical) between concepts are treated indifferently by the prototype, regardless of the semantics of the relationship. Thus, for example, a relationship such as "affects" had the same effect on the search expansion as a relationship such as "is associated with." No greater rigor was needed in establishing these relations.

The first ramification of the top concept of the ontology was made in a top-down manner by defining the analysis dimensions of the "outsourcing" domain, reflecting concepts from the value chain model (see item 4.2.1.1). In the bottom-up direction, terms were taken from the news, to ensure that there was no mismatch between the vocabulary of the ontology and the standard vocabulary (see paragraph 4.2.1.21). The selected terms (bottom-up) were complemented and grouped within the dimensions of the analysis (top-down), and hierarchies were created with the support of the IT outsourcing literature and the aid of the experts who participated in the construction of the ontology.

a. *Top-down construction of ontology*

The concepts of the second level of the ontology (just below the top concept "IT outsourcing"), called "dimensions analysis " here (in bold below), were derived from the Value Chain (Porter 1985), Value System (Porter 2008), and Five Forces Analysis (Porter 1979):

- From the Value Chain model:
  - Infrastructure was represented by the **Datacenters** dimension
  - Human Resource Management was represented by the **Human Resources** dimension
  - Technology Development was represented by the **Technological Resources** dimension
  - Operations were represented by the **Operation** dimension
  - Marketing and Sales were represented by **Motivators of outsourcing** dimension
  - Services were represented by the **Services** dimension. This dimension maps the portfolio of services from outsourcing providers.
- *From the Value System model:*
  - Suppliers were represented by the size Suppliers dimension in the ontology
  - Manufacturer was represented by the Outsourcing provider dimension in the ontology
  - Retail was represented by the Customers dimension in the ontology - the final link in the chain.
- *From the Five Forces Analysis model:*

The model of the Five Forces added no new dimensions to the ontology but was taken into consideration in the creation of the concepts below the dimensions. The following forces were considered: suppliers, potential entrants, buyers, and substitutes.

The top-down construction of the ontology resulted in its first two levels, as illustrated in Figure 4.

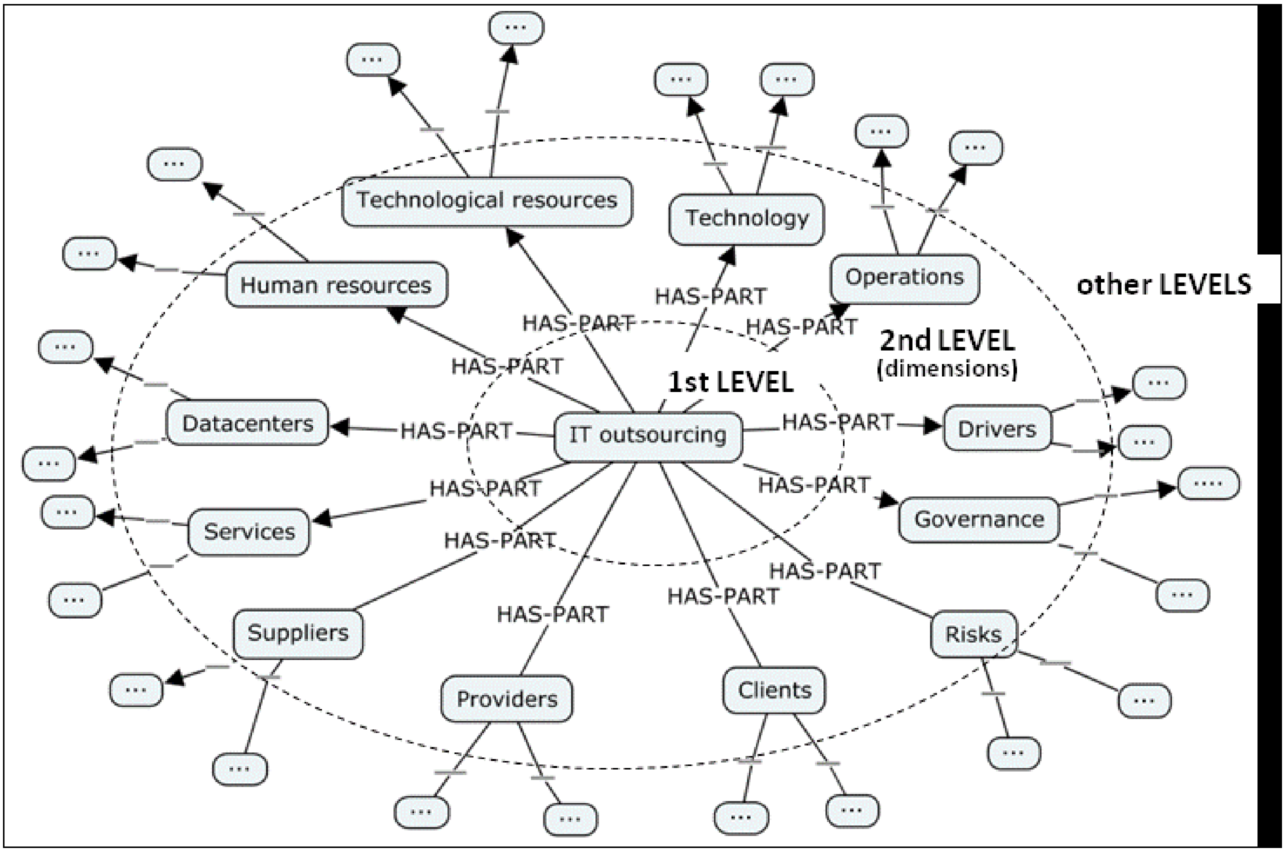

Figure 4 : Concepts of the first and second levels of the ontology

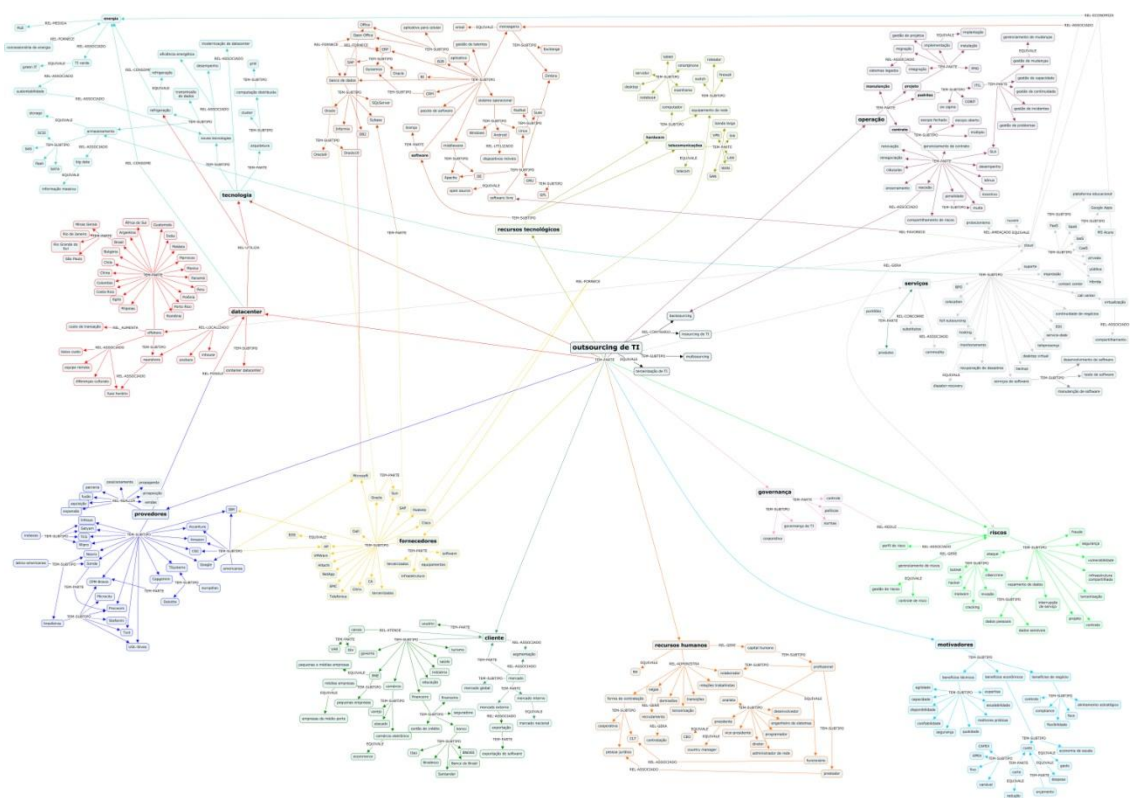

Figure 5 : Panoramic view of the complete ontology

b. Bottom-up construction of the ontology

In the bottom-up approach, terms for the ontology were manually extracted from a sample of 35 articles about IT outsourcing taken from leading national IT news sites in 2013, representing approximately 5% of the total.

c. Consolidation of top-down and bottom-up processes

Approximately 300 concepts and their hierarchical relationships, defined by the top-down and bottom-up processes, were designed in CMap Tools software. Based on the initial design, potential relationships between concepts from different hierarchies were analyzed for the definition of side (non hierarchical) relations. These relationships (approximately 400) were then added to the drawing.

This initial draft of the ontology was developed by the authors of this research, who are IT outsourcing experts. Two other experts in the field were then included in the process. Experts who currently occupy a range of different positions in the IT industry were selected to incorporate different perspectives.

The request for a design review was sent to the experts by email with an attachment containing the ontology in a PDF file. This was followed up by phone, at which points any questions about the request were discussed. Experts responded with suggestions by email and in telephone conversations, and suggestions were incorporated in the design of the ontology. The pilot ontology represented the consensus among the experts participating in the work. A panoramic view of the ontology is given in Figure 5, to convey an impression of the design layout.

c) System database

The ontology, graphically representing the concepts and their relationships, was exported to a text file in the form of RDF triples ([subject] [predicate] [object]), for example: "Oracle HAS-PART Sun" (Sun Microsystems has become an Oracle division after being acquired in 2009).

Below, we give samples of the triples found in this text-file related to the word "Oracle":

Subject Predicate Object

- "database HAS-SUBTYPE Oracle "
- "ERP HAS-SUBTYPE Oracle"
- "suppliers HAS-SUBTYPE Oracle "
- "Oracle REL-PROVIDES database"
- "Oracle REL-PROVIDES ERP "
- "Oracle REL-PROVIDES Open Office"
- "Oracle HAS-PART Sun"

The three-column ONTOLOGY table (SUBJECT, PREDICATE and OBJECT) was loaded with the fields from the exported text file.

The RDF triples of the ontology were then loaded into a single database table containing tree columns (SUBJECT, PREDICATE, and OBJECT), as shown in the ONTOLOGY table representation in Figure 5, following the vertical table model for representation and manipulation of ontologies (Dehainsala et al. 2007).

The other model tables were populated via execution of database scripts using the information in the Ontology table. The table CONCEPT (which contains all the ontological concepts) and the SUPERTYPE, SUBTYPE, ALL, PART, EQUIVALENT, and RELATED contained the related concepts, and the name of each table indicated the type of relationship. The table SUGGESTION, populated from those tables, would contain the expansion string for each concept of the ontology.

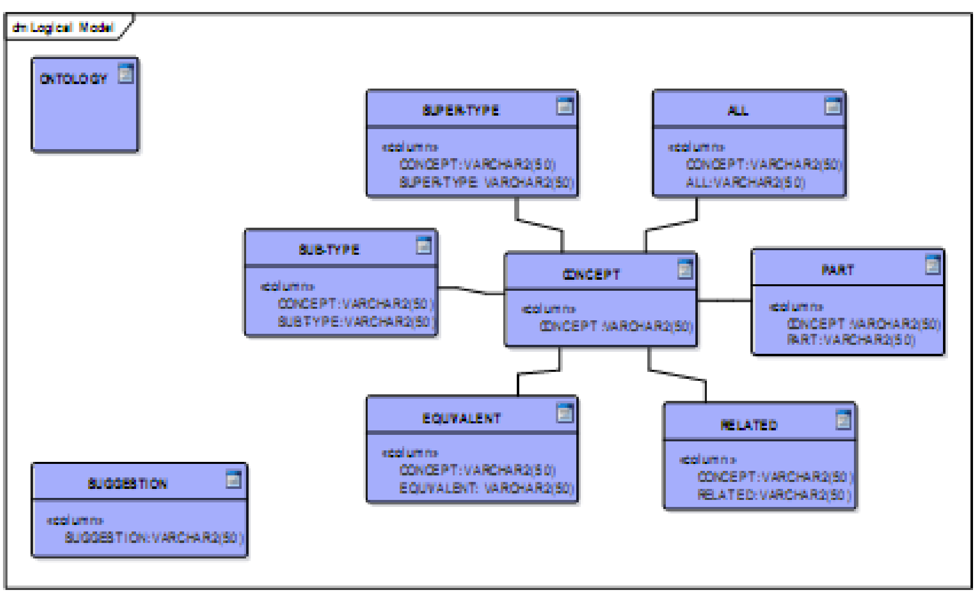

Figure 6 : System database tables

The script for the CONCEPTS table loaded concepts from both the SUBJECT and OBJECT column of the ONTOLOGY table, removing duplications.

For each concept stored in the CONCEPT table, the scripts for the peripheral tables SUPERTYPE, SUBTYPE, ALL, PART, EQUIVALENT, and RELATED loaded these tables with the associated concepts. The base concept then resided in the central table, and the concepts related to it in the peripheral tables, whose names indicated the type of relationship.

Finally, the script for the SUGGESTION table, based on the CONCEPT table and the peripheral tables, loaded the SUGGESTION table with groups of words suggested for search expansion.

The SUGGESTION table was then exported to the text file used by the TypingAid software.

Once TypingAid was configured to use the prepared text file, entering, for example, "Oracle" would generate the following strings as suggested replacements for the word "Oracle" (emulating a "self-complete"):

- "ORACLE DATABASE"
- "ORACLE ERP"
- "Oracle SUPPLIERS"
- "ORACLE OPEN OFFICE"
- "Sun ORACLE"

d) Construction of the interface

The interface for the search expansion system was built by integrating the MS Access database, which contained the ontology, with the Typingaid software and its auto-complete features. In this research, Typingaid was adapted to display a list of expressions to replace or complement each word typed into an input field of a standard search engine such as Google. Auto-complete requires the interface to anticipate the words or phrases that the user wants to type. In this study, the prediction was made at the semantic level. The system provided the keyword set that best defined the information needs of the user, based on the relationships between the concepts of the ontology, rather than on the most popular search terms, as in Google Suggest.

For each concept in the ontology, possible expansions were generated by adding concepts related to the original concept. The connection between the terms was done through the implicit logical operator "AND."

The terms suggested for expansion could be in uppercase or lowercase letters, depending on the relationship between the original term typed by the user and the terms suggested for expansion. This was designed to make explicit to the user (in case he is interested) whether the transit was from a more specific to a more general concept (moving up the hierarchy, a sort of drill-up), from a more general to a more specific concept (downward in the hierarchy, a sort of drill down), or to concepts in a nonhierarchical relationship with the original concept (side relationship - a kind of drill-across).

e) System Operation

The system operated in a manner similar to Google Suggest, which provides suggestions when using the Google search field.

For every term typed by the user, the system looked for concepts directly related to the term in the ontology (distance "1" in the networking concepts representing the ontology). The system then showed the user one or more strings composed by concatenating the original concept with related concepts, separated by "space" (corresponding to an implicit logical operator "AND" in the original configuration of the search engines). This guided the user to better contextualize the search term to obtain a more limited set of answers that were likely to contain the relevant elements. The following example illustrates the operation of the system: wider population. A total of 85 responses were obtained. The table below shows the average ratings for each item, evaluated on a seven-point Likert-type scale. Most evaluation scores were better than 4 ("neutral"), suggesting a good level of acceptance of the system.

Example: If the word "Oracle" is typed as the original search term, it will be expanded as indicated in Figure 7, according to the relations extracted from the ontology, leading the user to a disambiguation of terms.

Original Query	***Expanded Query***	**Remarks**
Oracle	• ORACLE DATABASE • ORACLE ERP • Oracle SUPPLIERS • ORACLE OPEN OFFICE • ORACLE Oracle10 • ORACLE Oracle9 • ORACLE Sun	• Oracle as database software • Oracle as ERP software • Oracle as supplier • Oracle as Open Office software (from Sun) • Subtype of Oracle database • Subtype de Oracle database • Sun as part of the Oracle company

Figure 7 : Expansions of the concept "Oracle"

f) System Tests

i. *Survey*

The system was trialed by a group of users who were asked to use the tool and complete the evaluation questionnaire. In the test, the users installed the TypingAid software with a sentences file preloaded with the terms of the ontology. They then performed searches related to "IT outsource" using the Google search tool but taking the suggested terms from the ontology (through TypingAid) instead of the suggestions made by Google itself. The users then completed the adapted TAM3 questionnaire, to measure their acceptance level.

The following topics were suggested: outsourcing risks, the Oracle outsourcing market, cloud projects, professional experts in outsourcing, available services for outsourcing, outsourcing providers, and technologies used in outsourcing.

The sample of 297 participants was recruited by email by using convenience sampling. The population comprised professionals and researchers in the IT field who were either members of an Information Systems study group at the university or professional contacts of the research team working in IT areas of business (for example outsourcing, project management, software development, or banking IT departments). The results obtained therefore cannot be generalized to the wider population. A total of 85 responses were obtained.

The table below shows the average ratings for each item, evaluated on a seven-point Likert-type scale. Most evaluation scores were better than 4 ("neutral"), suggesting a good level of acceptance of the system.

Table 1: Average scores from the survey

Group	Statement	Average of the answers
Perceived Usefulness	1. Using the system improves my performance in my job.	5,69
	2. Using the system in my job increases my productivity.	5,72
	3. Using the system enhances my effectiveness in my job.	5,54
	4. I find the system to be useful in my job.	5,69
Perceived Ease of Use	5. My interaction with the system is clear and understandable.	5,33
	6. Interacting with the system does not require a lot of my mental effort.	5,82
	7. I find the system to be easy to use.	5,82
	8. I find it easy to get the system to do what I want it to do.	5,48
Computer Self-Efficacy	9. I could complete the job using a software package if there was no one around to tell me what to do as I go.	4,88
	10. I could complete the job using a software package if someone showed me how to do it first.	5,08
	11. I could complete the job using a software package if I had used similar packages before this one to do the same job.	4,88
Perceptions of External Control	12. I have control over using the system.	5,24
	13. I have the resources necessary to use the system.	6,33
	14. Given the resources, opportunities and knowledge it takes to use the system, it would be easy for me to use the system.	6,24
	15. The system is compatible with other systems I use.	5,66
Perceived Enjoyment	16. I find using the system to be enjoyable.	5,34
Output Quality	17. The quality of the output I get from the system is high.	5,33
	18. I have no problem with the quality of the system's output.	5,42
	19. I rate the results from the system to be excellent.	5,2
	20. I have no difficulty telling others about the results of using the system.	5,91
	21. I believe I could communicate to others the consequences of	

Result Demonstrability	using the system.	5,84
	22. The results of using the system are apparent to me.	5,8
	23. I would have difficulty explaining why using the system may or may not be beneficial.	5,86
Behavioral Intention	24. Assuming I had access to the system, I intend to use it.	5,78
	General Average	**5,58**

ii. *Factor analysis*

Factor Analysis generated six key factors (F1–F6), representing 24 variables (V1–V24) corresponding to the items in the questionnaire. The original TAM3 model had eight factors, whereas the adapted version in this research found only six factors. This was possibly because of the smaller number of items in the adapted questionnaire. The reduced number of variables also reduced two of the original TAM3 factors to a single variable each (factors: "nice use Perception" and "behavioral Intent"). These variables would be isolated in the original factors and were then associated with other factors in the factor analysis. Aside from this simplification, the factors coincide with the conceptual model of TAM3, making it consistent with our survey.

The marks in the table below indicate the factors (columns 1–6) to which the variables (V1–V24 lines) are most strongly associated with, as they show the biggest factor loads:

Table 2 : Factor loads matrix

	Component					
	1	2	3	4	5	6
V1	,098	,162	,390	,704	,125	,171
V2	,121	,106	,273	,798	,124	,001
V3	,287	,177	,029	,731	,209	,127
V4	,064	,282	,231	,764	,066	,055
V5	,823	,244	,173	,107	,088	,016
V6	,853	,196	,214	,085	,046	-,017
V7	,805	,321	,013	,215	,024	,161
V8	,779	,323	,192	,148	,076	,065
V9	,025	-,004	,030	,081	,925	-,090
V10	,072	,039	,016	,152	,950	,043
V11	,147	,119	,153	,141	,870	,102
V12	,439	,544	,140	,150	,113	,060
V13	,027	-,021	,105	,131	-,020	,883
V14	,544	,078	,438	,059	,306	,251
V15	,203	,361	,178	,082	,075	,673
V16	,453	,638	,237	,196	,041	,139
V17	,283	,755	,229	,330	,067	,060
V18	,354	,797	,270	,023	,050	,054
V19	,260	,777	,186	,373	,008	,097
V20	,322	,095	,737	,159	,023	,080
V21	,081	,165	,845	,194	,163	,014
V22	,065	,408	,682	,195	,045	,191
V23	,338	,194	,677	,309	-,055	,106
V24	,120	,421	,585	,362	,065	,119

V. Discussion

The user evaluation of the prototype suggests that an interactive expansion tool for internet searches based on an ontology of the target business domain helps users refine their searches.

The business domain ontology was built manually from business knowledge, with a vocabulary alignment based on a sample of news, and incorporating competitive models. This has shown promise as a tool for the selection of news, regardless of the fact that news items are dynamically changing, which presents an extra challenge for the alignment of ontology terms and news terms.

Although the results should not be generalized for the population represented in the survey, the proposed system proved a useful tool for mitigating information overload in internet searches. Adding structure to unstructured information gave users greater control over the information retrieved from online news databases and helped them to narrow down their searches.

Finally, we revisit below the research questions and objectives of the study, to judge the contribution of the research. As this was an exploratory study, the findings were not tested statistically. However, they provide material for future research.

- Q1–Is a manually constructed business domain ontology incorporating competitive models useful as a resource for news selection (dynamic database)?

This question was addressed by the specific objectives O1 and O3.

Contributions of the research:

- We applied information retrieval based on ontological concepts with a volatile textual basis, whereas previous works have generally dealt with static or quasi-static textual bases.
- We applied information retrieval using ontology as an information gathering tool for business domain competitive intelligence, whereas previous works have mostly targeted textual databases (for example, collections of libraries) which are unrepresentative of the market news used by businesses.
- Ontology development was based on specific business knowledge, whereas previous works have used ready-made ontologies, or used allegorical ontology unrepresentative of the real situation of business domains.
- Q2–Does the use of ontological relationships to expand the search terms increase the selectivity of the information retrieved?

This question was addressed by all three specific objectives O1 to O3.

Contributions of the research:

- We applied the concept of facets, widely used in structured databases, to the retrieval of textual information through the expansion of search terms by ontological side relations.
- Q3 - How can the business domain ontology be used to reduce information overload?

This question was addressed the the specific objectives O2 and O3.

Contributions of the research:

- We proposed a solution software architecture, based on established models, taking widely used search tools and adding features that tackle the problem of information overload.
- We created a functional prototype representing a class of solutions to a class of problems, based on Design Research methodology. The proposed architecture can be applied to equivalent problems in other areas of business, as well the "IT outsourcing" example used in this research.

References References Referencias

1. Bower, G.H. & Hilgard, E.R., 1981. *Theories of Learning*, Prentice-Hall. Available at: http://www.brookes.ac.uk/services/ocsd/2_learntch/theories.html#sensory.
2. Brennan, L.L., 2006. The Scientific Management of Information Overload. *Journal of Business and Management*, (1997), pp.121–135.
3. Cao, T.D., 2006. *Exploitation du web sémantique pour la veille technologique*. UNIVERSITE de Nice-Sophia Antipolis.
4. Chung, W., Chen, H. & Jr, J.A.Y.F.N., 2005. A Visual Framework for Knowledge Discovery on the Web : An Empirical Study of Business Intelligence Exploration. *Journal of Management Information Systems*, 21(4), pp.57–84.
5. Cohen, W.M. & Levinthal, D.A., 1990. Absorptive Capacity : A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1), pp.128–152.
6. Crossan, M.M. et al., 1999. An Organizational Learning Framework: From Intuition to Institution. *The Academy of Management Review*, 24(3), pp.522–537.
7. Davis, F.D., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology T. Ball & R. B. Jones, eds. *MIS Quarterly*, 13(3), pp.319–340. Available at: http://www.jstor.org/stable/249008?origin=crossref.
8. Davis, N., 2011. 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*, pp.45–50.
9. Dehainsala, H., Pierra, G. & Bellatreche, L., 2007. OntoDB : An Ontology-Based Database for Data Intensive Applications. In *Proc . of Database Systems for Advanced Applications (DASFAA ' 2007)*.
10. Denton, D.K. & Richardson, P., 2012. Using Intranets to Reduce Information Overload. *Journal of Strategic Innovation and Sustainability*, 7(3), pp.84–95.
11. Eppler, M.J. & Mengis, J., 2004. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society*, 20(5), pp.325–344. Available at: http://www.tandfonline.com/doi/abs/10.1080/01972240490507974 [Accessed October 28, 2012].
12. Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing N. Guarino & R. Poli, eds. *International Journal of*

Human-Computer Studies, 43(5-6), pp.907–928. Available at: http://linkinghub. elsevier.com/retrieve /doi/10.1006/ijhc.1995.1081.
13. Gulla, J.A., Borch, H.O. & Ingvaldsen, J.E., 2007. Ontology Learning for Search Applications. In *OTM'07 Proceedings of the 2007 OTM Confederated international conference on On the move to meaningful internet systems: CoopIS, DOA, ODBASE, GADA, and I*. pp. 1050–1062.
14. Hwang, W. & Salvendy, G., 2005. Effects of an ontology display with history representation on organizational memory information systems. *Ergonomics*, 48(7), pp.838–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16076741[Accessed November 10, 2012].
15. Jenkin, T.A., 2008. *USING INFORMATION TECHNOLOGY TO SUPPORT THE DISCOVERY OF NOVEL KNOWLEDGE IN ORGANIZATIONS*. Queen's University.
16. Jonker, C.M., Riemsdijk, M.B. van & Vermeulen, B., 2010. Shared Mental Models - A Conceptual Analysis. In *Proceedings of the 9th International Workshop on Coordination, Organization, Institutions and Norms in Multi-Agent Systems*.
17. Kudryavtsev, D., 2006. Mapping ontologies and contexts : from theory to a case study . In *C&O-2006*. Riva del Garda, Italy, pp. 9–11.
18. Kuhlthau, C.C., 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5), pp.361–371. Available at: http://doi.wiley.com/10.1002/(SICI) 1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#.
19. Lankton, N.K., Speier, C. & Wilson, E.V., 2012. Internet-based knowledge acquisition: Task complexity and performance. *Decision Support Systems*, 53(1), pp.55–65. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167923611002417 [Accessed November 10, 2012].
20. Li, T., 2011. An Investigation and Analysis of Information Overload in Manager's Work. *iBusiness*, 03(01), pp.49–52. Available at: http://www.scirp.org/journal/PaperDownload.aspx?DOI=10.4236/ib.2011.31008 [Accessed November 10, 2012].
21. Li, Y. et al., 2012. A two-stage decision model for information filtering. *Decision Support Systems*, 52(3), pp.706–716. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167923611002090 [Accessed November 10, 2012].
22. Maes, P., 1994. Agents that reduce work and information overload. *Communications of the ACM*, 37(7), pp.30–40. Available at: http://portal.acm.org/citation.cfm?doid=176789.176792.
23. Marshall, B. et al., 2004. EBizPort: Collecting and analyzing business intelligence information. *Journal of the American Society for Information Science and Technology*, 55(10), pp.873–891. Available at: http://doi.wiley.com/10.1002/asi.20037 [Accessed November 10, 2012].
24. Mizoguchi, R., 2003. Tutorial on ontological engineering. *New Generation Computing*, 21(4), pp.363–364. Available at: http://www.springerlink.com/index/10.1007/BF03037310.
25. Noy, N.F. & Mcguinness, D.L., 2001. *Ontology Development 101 : A Guide to Creating Your First Ontology* S. A. McIlraith, D. Plexousakis, & F. Harmelen, eds., Citeseer. Available at: http://bmir.stanford. edu/file_asset/index.php/108/BMIR-2001-0880.pdf.
26. Ong, T.-H. et al., 2005. Newsmap: a knowledge map for online news. *Decision Support Systems*, 39(4), pp.583–597. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167923604000545 [Accessed November 10, 2012].
27. Porter, M.E., 1985. *Competitive Advantage: Creating and Sustaining Superior Performance* Anonymous, ed., Free Press. Available at: http://books.google.com/books?id=_Fu1AAAAIAAJ&pgis=1.
28. Porter, M.E., 1979. How competitive forces shape strategy. *Harvard Business Review*, 57(2), pp.137–146. Available at: http://www.ncbi.nlm. nih.gov /pubmed/18271320.
29. Porter, M.E., 2008. *On Competition, Updated and Expanded Edition*, Harvard Business School Press. Available at: http://harvardbusinessonline. hbsp.harvard.edu/relay.jhtml?name=itemdetail&id=2696.
30. Silver, M.S., 1991. Decisional Guidance for Computer-Based Decision Support. *MIS Quarterly*, 15(1), pp.105–122. Available at: http://www.jstor.org/stable/249441?origin=crossref.
31. Tate, C.C., 2008. *Using Visualization Tools to Mitigate Information Overload on the Internet*. Georgetown University.
32. Uschold, M. & Gruninger, M., 1996. Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11(2), pp.93–136. Available at: http://journals.cambridge.org/abstract_S0269888900007797.
33. Vaishnavi, V. & Kuechler, B., 2004. Design Science Research in Information Systems Overview of Design Science Research. *Association for Inormation Systems*, (1978), pp.1–16. Available at: http://desrist.org/desrist.
34. Venkatesh, V. & Bala, H., 2008. Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, 39(2), pp.273–315. Available at: http://doi.wiley. com/10.1111/j.1540-5915.2008. 00192.x.
35. Village, D., 2000. Cyberspace 2000 : Dealing with Information Overload. *COMMUNICATIONS OF THE ACM*, 40(2), pp.19–24.

This page is intentionally left blank