Expert Systems with Applications 38 (2011) 15565–15572

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Financial news semantic search engine

Eduardo Lupiani-Ruiz, Ignacio García-Manotas, Rafael Valencia-García *, Francisco García-Sánchez, Dagoberto Castellanos-Nieves, Jesualdo Tomás Fernández-Breis, Juan Bosco Camón-Herrero

Department of Computer Science and Systems, University of Murcia, Spain

ARTICLE INFO

Keywords:
Semantic search engine
Semantic Web
Ontologies
Ontology population

ABSTRACT

An increasingly large amount of financial information available in a number of heterogeneous business sources implies that the traditional methods of analysis are no longer applicable. These financial data sources are characterized by the use of disparate data models and their unstructured content with implicit knowledge. In addition, the most up-to-date financial information typically resides in the vast amount of financial-related news that brokers take into account when investing. As Semantic Technologies mature, they provide a consistent and reliable basis for the development of superior, more precise mechanisms to deal with heterogeneous data. In this paper, we present a financial news semantic search engine based on Semantic Web technologies. The search engine is accompanied by an ontology population tool that assists in keeping the financial ontology up-to-date. In addition, a further module has been developed that is capable of crawling the Web in search of financial news and annotating it with knowledge entities from the financial ontology that match with the contents of the news. Our contribution is an overall solution based on a fully fledged architecture that has been validated in a use case scenario for the Spanish stock exchange.

1. Introduction

The stock market exchange market involves a large number of business and companies with a tremendous economic impact on our society. The need to manage financial data has been coming into increasingly sharp focus for some time. Years ago, these data sat in silos attached to specific applications in banks and financial companies. Then, the Web entered the arena, making diverse data sets available across applications, departments and other financial entities. However, throughout the course of these developments, a particular underlying problem has remained unsolved: data reside in thousands of incompatible formats and cannot be systematically managed, integrated, unified or cleansed. To make matters worse, this incompatibility affects not only the use of different data technologies (for example, the different relational databases available (in existence)), but also its incompatibility in terms of semantics. Despite the complexity of the domain, financial companies and end-users deem as necessary a fully fledged integrated approach to cope with the ever-increasing volume of information outperforming current approaches such as Yahoo Finance.

Consequently, more accurate and powerful information management strategies are required, and this community needs tools with the ability to acquire, communicate, and disseminate business information which is vital for investor and management decision making (Pinsker & Li, 2008).

The enormous success of search engines has demonstrated the value of crawling and indexing web resources. The working hypothesis of the current project is that search engines may help the financial sector to deal with those new challenges. Our proposal is not a traditional search engine, but a semantic one. Our solution is inspired by the Semantic Web (SW) approach, whose main challenge is to enable better machine information processing by structuring web documents to make them more easily machine readable. These technologies can help provide users with the proper support to take advantage of the huge amount of information available on the Web. In the Semantic Web, knowledge is represented by means of ontologies, which are viewed in this work as a formal specification of a domain knowledge conceptualization (Studer, Benjamins, & Fensel, 1998).

This semantic search engine has been specially designed for dealing with financial news, and the developed software application makes use of a financial ontology for the semantic indexing and annotation of natural language documents.

The remainder of this paper is organized as follows. In Section 2, the fundamentals of knowledge representation and ontology population are briefly described. Different semantic search tools and

* Corresponding author. Fax: +34 868884151.
E-mail addresses: elupiani@um.es (E. Lupiani-Ruiz), ignacio.g.m@um.es (I. García-Manotas), valencia@um.es (R. Valencia-García), frgarcia@um.es (F. García-Sánchez), dcastellanos@um.es (D. Castellanos-Nieves), jfernand@um.es (J.T. Fernández-Breis), jbcamon@um.es (J.B. Camón-Herrero).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.06.003

related work are also analyzed in this section. The proposed architecture of the financial news semantic search engine is shown in Section 3. The validation in the Spanish stock market is reported in Section 4. Finally, our conclusions are outlined in Section 5.

2. Background

2.1. Towards semantic search engines

Semantic search engines differ from traditional ones in two main key respects (Esmaili & Abolhassani, 2006): (1) they use a logical framework that makes more intelligent retrieval possible; (2) the management of complex semantic relationships makes the meta-data maintenance harder though it favors more sophisticated ranking mechanisms.

The amount of information published on the Web is enormous and grows with each passing day. The traditional search engines, as Google or Yahoo are constantly building indexes so they create ties between words and documents, so that when a user submits a query, the search engines return its related documents. The result is a large set of documents, in most cases "shorted" (i.e. sorted and shortened) by an algorithm such as Page Rank (Langville & Meyer, 2006). The engine does not understand the meaning of the query, so the results include all the possible alternatives; for example if the user types the word "Enterprise", the result should be a set with documents related to "Enterprise" and those related with the spacecraft "Enterprise" from Star Trek. Furthermore, the result would not include documents containing the word "Company", which is a synonym of "Enterprise", or "Empresa" from its translation into Spanish.

The quality of the results can be improved by categorizing these using ontologies. For instance, the search engine would harvest news pages related to companies and it could classify them into a structure of types of companies according to the economical activity, such as energy companies or financial companies. This classification would also make it possible to easily filter the results by economic activity. In our opinion, the semantic search engines obtain better results because they understand the query meaning. Consequently their accuracy is higher.

A fundamental prerequisite of the Semantic Web is the existence of large amounts of meaningfully interlinked RDF/OWL data on the Web (Bizer, Heath, Ayers, & Raimond, 2007). RDF is a data-model for information representation and OWL is the Web Ontology Language used for publishing and sharing explicit and common descriptions of domain knowledge, and providing support for efficient knowledge management. Both representations are W3C recommendations for modeling ontologies in the SW.

There are four categories of semantic search engines according to their user interface (Lei, Uren, & Motta, 2006): (i) form-based, engines which provide sophisticated web forms that allow users to specify queries by selecting ontologies, classes, properties, and values (e.g., RKBExplorer, mspace); (ii) RDF-based query languages, which provide sophisticated querying languages to support semantic search (e.g., DBpedia Snorql, DBpedia Movile, Dbin, SPARQLBot, FOAF@QDOS, ⟨*sameAs*\⟩; (iii) semantic-based keywords, which increase the performance of traditional keyword search techniques by making use of available semantic data (e.g., MultimediaN, FreeBase, Hakia, SenseBot, Powerset, DeepDybe, Cognition, BBC Programmes, BBC Music, Revyu, Foafin); and (iv) question answering tools, which exploit available semantic markup to answer questions asked in natural language format (e.g., LEXXE).

The system that we present in this paper is a semantic-based keyword search engine, because the system uses a financial ontology as kernel of its processing system. Another system called GoWeb (Dietze & Shoreder, 2009) has demonstrated that this kind of semantic searchers have a success rate of up to 79% in the gene domain.

2.2. Ontologies and finances

The formal semantics underlying ontology languages enables the automatic processing of the information and allows the use of semantic reasoners to infer new knowledge. Ontologies provide a formal, structured knowledge representation, with the advantage of being reusable and shareable. Ontologies provide a common vocabulary for a domain and define with different levels of formality the meaning of the terms and the relations between them. Ontologies also provide the meaning and facilitates the efficient retrieval of contents and information (Guo & Zhang, 2009), as well as improving crawling (Yang, 2009). Knowledge in ontologies is mainly formalized using five kinds of components: classes, relations, functions, axioms and instances (Gruber, 1993). Classes in the ontology are usually organized into taxonomies. Sometimes, the definition of ontologies has been diluted, in the sense that taxonomies are considered to be full ontologies (Studer et al., 1998). In this work, the Ontology Web Language (OWL),[1] which is the Semantic Web standard language, has been used to represent the knowledge extracted from texts.

Semantic Technologies are currently achieving a certain degree of maturity. They provide a consistent and reliable basis to face the challenges named before and aiming at a fine-grained approach for organization, manipulation and visualization of the financial data (Castells, Foncillas, Lara, Rico, & Alonso, 2004). In the last few years, several finances-related ontologies have been developed. The ontology TOVE (Toronto Virtual Enterprise) (Fox, Barbuceanu, Gruninger, & Lin ,1997), developed by the Enterprise Integration Laboratory from Toronto University, describes a standard organization company as their processes. BORO (Business Object Reference Ontology) is intended to be suitable as a basis for facilitating, among other things, the semantic interoperability of enterprises' operational systems (Partridge & Stefanova, 2001) The consortium DIP (Data Information and Process Integration) has developed an ontology for the financial domain which focuses on the description of Semantic Web services in the stock market domain (Alonso et al., 2005). The XBRL Ontology Specification Group developed a set of ontologies for describing financial and economical data in RDF for sharing and interchanging data. This ontology is becoming an open standard means of electronically communicating information among businesses, banks, and regulators (XBRL International., 2009).

2.3. Ontology population

One of the main problems with ontologies is their construction. Several methodologies have been developed to assist in building ontologies. Yet, the manual construction of ontologies is still considered a major bottleneck. Several research studies are being carried out with the aim finding mechanisms to automate ontology building. In this context, two major categories can be distinguished: ontology learning (Maedche & Staab, 2004) and ontology population (Ruiz-Martínez et al., 2008a, 2008b).

Ontology learning (also named ontology generation or ontology extraction) is a knowledge acquisition activity that relies on (semi-) automatic methods to transform unstructured (e.g., corpora), semi-structured (e.g., folksonomies, html pages, etc.) and structured data sources (e.g., databases) into conceptual structures. Some ontology learning approaches, such as TERMINAE (Aussenac-Gilles, Despres,

[1] http://www.w3.org/TR/owl-features/.

& Szulman, 2008), provide guidance to conceptualization from natural language text integrating functions for linguistic analyses and conceptual modeling. Ontology population, on the other hand, is a knowledge acquisition activity that relies on (semi-) automatic methods to transform unstructured, semi-structured and structured data sources into instance data. In other words, whereas ontology learning deals with the acquisition of new concepts and relations with the consequence of changing the definition of the ontology itself, ontology population pursues the extraction and classification of instances of the concepts and relationships defined in the ontology. The instantiation of the ontology with new knowledge is a significant step towards the provision of valuable ontology-based knowledge services.

We can distinguish two types of ontology population: (i) from free text, and (ii) from semi-structured documents such as XML, HTML, RSS, etc. Concretely in this work we have developed a semi-automatic method for ontology population from semi-structured texts.

Most Web content is provided as semi-structured or unstructured HTML documents. Wrapping information from HTML tables has received much attention in last few years (Sugibuchi & Tanaka, 2005), and this information is usually represented by means databases (Pan, Raposo, Alvarez, Carneiro, & Bellas, 2007) or is transformed into semantic annotations (Tao & Embley, 2009).

There are different approaches for populating ontologies from semi-structured or unstructured HTML documents. For example, in the work presented in (Seong-Bae et al., 2008) an ontology is populated using RDF triples obtained from HTML tables. HTML documents are obtained from a Web crawler and HTML tables are processed using wrappers based on predefined patterns. The Levenshtein distance is used to identify which properties of the table are equivalent to the properties of concepts in the ontology, so they do not use any semantic information.

3. News search engine architecture

The proposed architecture for the financial news semantic search engine is consists of three main modules (see Fig. 1): the financial ontology, the ontology population module, and the ontology-based search engine module. We will now turn to a detailed explanation of these models.

3.1. Financial ontology

We have developed a financial OWL ontology based on the available ontologies described in Section 2.2. This ontology has 247 classes, 86 subclass axioms, 34 data type properties, 38 object properties and 87 restrictions.

The ontology covers four main financial concepts (see Fig. 2):

- **Financial market**: A mechanism that allows people to easily buy and sell financial assets such us stocks, commodities, currencies, etc. The main stock markets such as Nasdaq, London Stock Exchange or Madrid Stock Exchange have been modeled in the ontology as subclasses of Stock_Market.
- **Financial Intermediary**: The entities that typically invest in the financial markets. Examples of such entities are banks, insurance companies, brokers and financial advisers.
- **Asset**: Everything that has a value and can be the object of an investment, such as stock market indexes, commodities, companies, currencies, etc. So, for instance, enterprises such

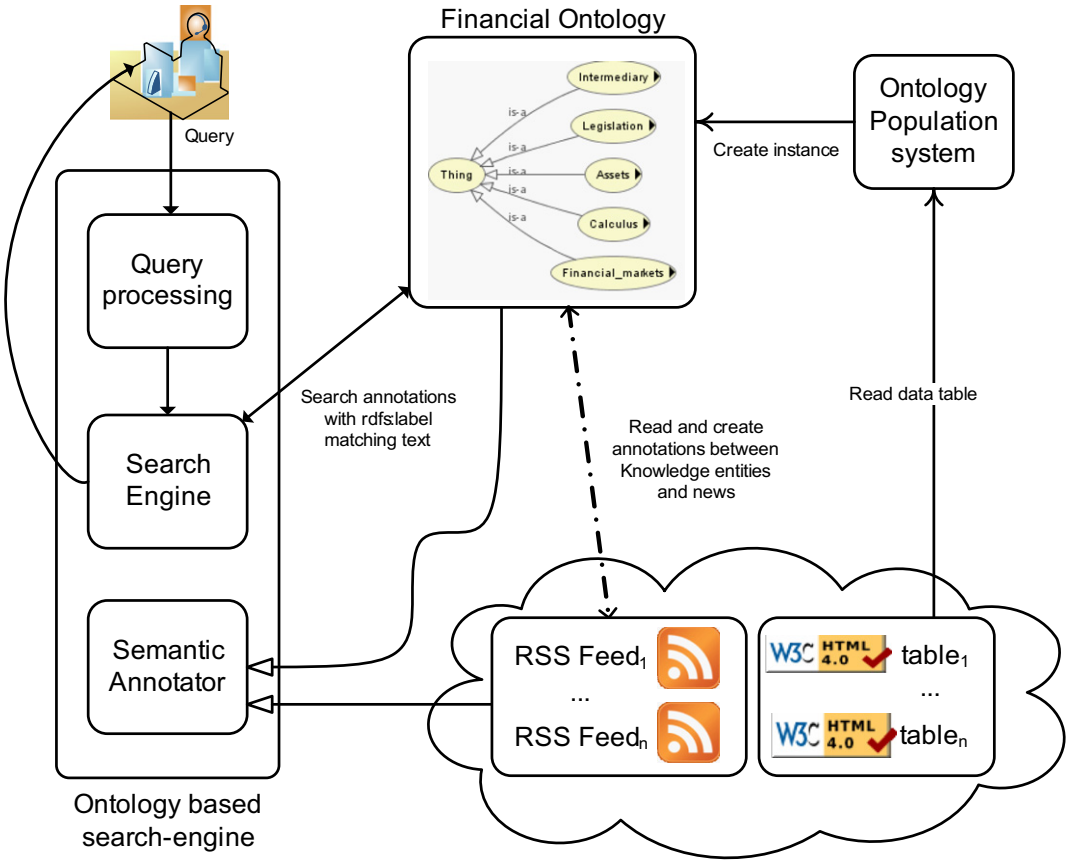

Fig. 1. System architecture.

Fig. 2. Excerpt of the financial ontology.

as General Electric or Microsoft are individuals of the class Company and currencies such as US dollar or Euro are individuals of Currency.

- **Legislation**: The entities that supervise the stock markets (e.g., the Federal Reserve or the International Monetary Fund), and the regulation and laws that can be applied to the financial domain.

The Jena Semantic Web Framework has been used for obtaining relational persistence, due to it its simplicity and support for fast software development. In particular, MySQL has been the database management system used in this work.

3.2. Ontology population from heterogeneous data sources

The ontology population module gathers knowledge from semi-structured and non-structured texts. The ultimate goal of our approach is to populate the financial ontology with all the relevant information identified. The populated ontology will then serve as the keystone component for an up-to-date, knowledge-based search engine. The architecture of the proposed ontology population system is based on previous works (García-Manotas, Lupiani, García-Sánchez, & Valencia-García, 2010) and is shown in Fig. 3. It is composed of three main components: (i) a set of selection systems (SIS), (ii) the "Selection and Converter System" (TSiR) module, and (iii) the "Massive Population Algorithm" (Mpa) module. The input of the system is represented at the top of Fig. 3. It consists of a collection of Web-available information resources. The tool has been designed to support both semi-structured and non-structured texts. The output of this module is a number of ontology instances that are stored in the repository. The storage submodule is shown at the bottom of Fig. 3.

In a nutshell, the system works as follows. Semi-structured or non-structured data sources available on the Internet are parsed to extract the information that can be gathered from the text. Currently, only semi-structured elements from HTML- and RSS-formatted documents are supported by the system. However, the platform can be easily extended to support other kind of resources.

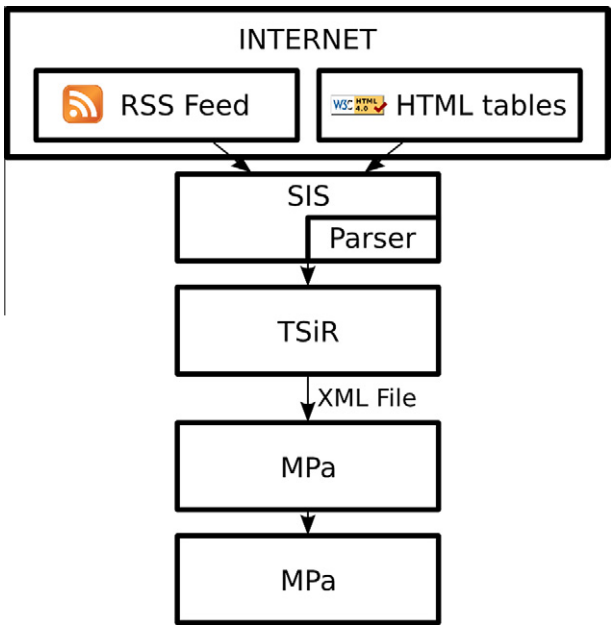

Fig. 3. Ontology population system.

In particular, the tables contained in the HTML documents and text included in RSS documents constitutes the semi-structured information used in this system. Users are shown the parts of the semi-structured texts identified by the parser.

Users must choose which of the found elements are relevant and have to be stored in the knowledge base. Users have to set up two parameters: (i) a set of substitution or transformation rules, which will be used by the TsiR module to transform the information into the appropriate format, and, optionally (ii) the set of ontological concepts that are related to the information elements to be gathered from the source semi-structured text. This latter optional parameter aims to improve the efficiency and accuracy of the Mpa module. Once users have indicated the tables of the resources in which they are interested, the TSIR module transforms the tables into an internal format in XML. For this purpose, the aforementioned user-defined transformation rules are applied. During this process, the position of the information in the tables is taken into account to form groups. Each group is represented in the form of tuples ⟨attribute, literal⟩. The XML file produced by the TsiR and the set of ontology concepts pointed out by the user are the input of the Mpa module. With this information, Mpa generates the correspondences between the data in the semi-structured texts and the concepts in the ontology. Finally, the new discovered ontology instances are stored in the knowledge base.

This module explores the leading pages concerning the stock market exchange to populate the ontology with financial information. For instance, Yahoo! Finance[2] is a popular portal which offers stock information for different index as NASDAQ, S&P 500, FTSE 100, Nikkei 225, Hang Seng, etc. The crawling system reads the tables of components and their information, immediately storing it in the corresponding knowledge entity of the ontology or creating a new instance if the information is actually new, as a new company in the index.

Fig. 4 rates the Hang Seng components from Yahoo! Finance. The population system will process each row of the table and will create a new instance of the company "Cheung Kong" in case it does not exist, and other information such as "Last Trade", "Change" and "Volume".

3.3. Ontology-based search engine

The search engine has been divided into three modules (see Fig. 1): (i) Semantic Annotator, (ii) Query Processing, and (iii) Search Engine.

Semantic Annotator: In this module, news from RSS feeds and Web Pages are semantically annotated by the domain ontology. The process that takes place during the semantic annotation is as follows. First, the most important linguistic expressions are identified using statistical approaches based on term extraction methods. Then, for each linguistic expression, the system tries to determine whether the expression under question is an individual of any of the classes of the domain ontology. Next, the system retrieves all the annotated knowledge that is situated next to the current linguistic expression in the text, and tries to create fully filled annotations with this knowledge. This process has been implemented using GATE.[3] Fig. 5 illustrates an example of the annotation process of financial news using GATE.

Query Processing: Users can query the system through natural language queries based on previous works of our research group (Ruiz-Martínez et al., 2008a, 2008b). For this, four main steps are carried out. First, a POS-Tagging process is performed. This allows the system to identify the grammar category of each word in the sentence and removes the non-content words. Then, the system identifies the lemma of each word by means of a lemmatizing process. A chunking and name entity recognition process is performed in order to obtain the focus of the query. Finally, the synonyms related to the financial domain are listed.

Search Engine: In OWL-based ontologies, the *rdfs:label* is an instance of *rdf:property* that may be used to provide a human-readable version of a resource name. In this work, all the resources in the ontology have been annotated with the label descriptor. The main objective of this module is to identify the financial news that is related to the expanded query obtained from the *Query Processing* module and to sort these results. The sorting function is based on semantic similarity functions (Maedche & Staab, 2002) in order to obtain the most related financial news from the user query.

The system is constantly crawling news information from HTML and RSS feeds with the Annotator Module, creating semantic annotations for the news pages. If no annotations have been created for news then the news is not stored in the database. The financial ontology is continuously populated, so the annotation process is executed periodically.

The user queries the system through the Query Processing module and the semantic information of the query is then passed to the *Search Engine* which is in charge of obtaining the news that is semantically related to the knowledge entities obtained from the query.

For example, let us suppose that the ontology contains the taxonomy presented in Fig. 5. There are two kinds of companies, namely, "Aviation" and "Energy". Each of these classes contains a set of individuals such us "GE Aviation" and "GE energy" respectively. If the user is searching for news of "General Electric Aviation" (a.k.a. GE) the system will return all news annotated with the individual GE Aviation, but news related to other *Aviation* companies could be relevant to the user, so the system shows news about Boeing and Airbus. If the user queries the system for "Energy companies", then the result will include all the news that contains the concept "Energy company" thereby obtaining all news related to the "GE Energy", "Texaco" and "Shell" companies. Furthermore, if the query is such a general word as "Company", the user is given the possibility of filtering the results according to the subclasses of

[2] http://www.finance.yahoo.com.

[3] General Architecture for Text Engineering http://www.gate.ac.uk/.

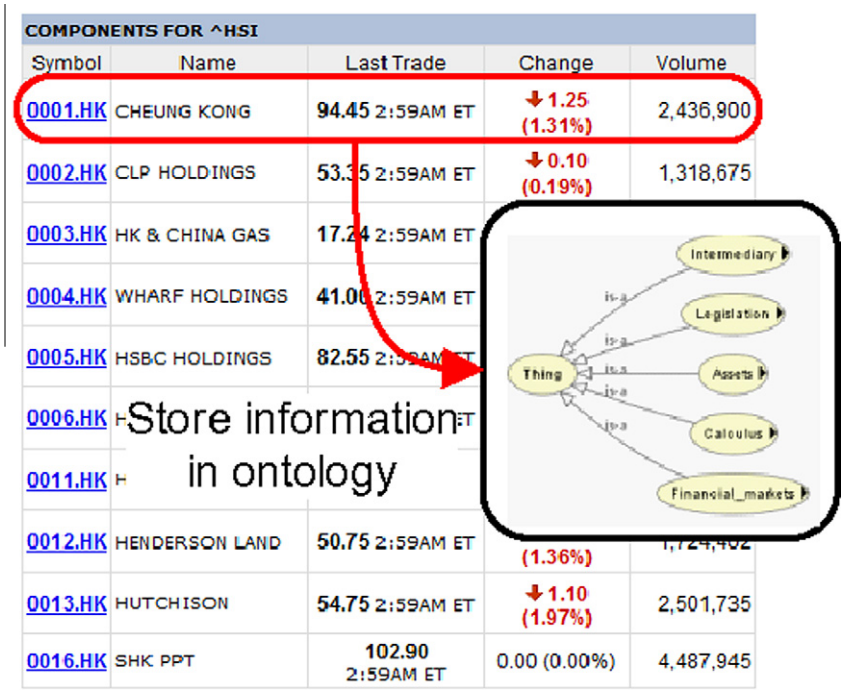

Fig. 4. Ontology population from HTML tables.

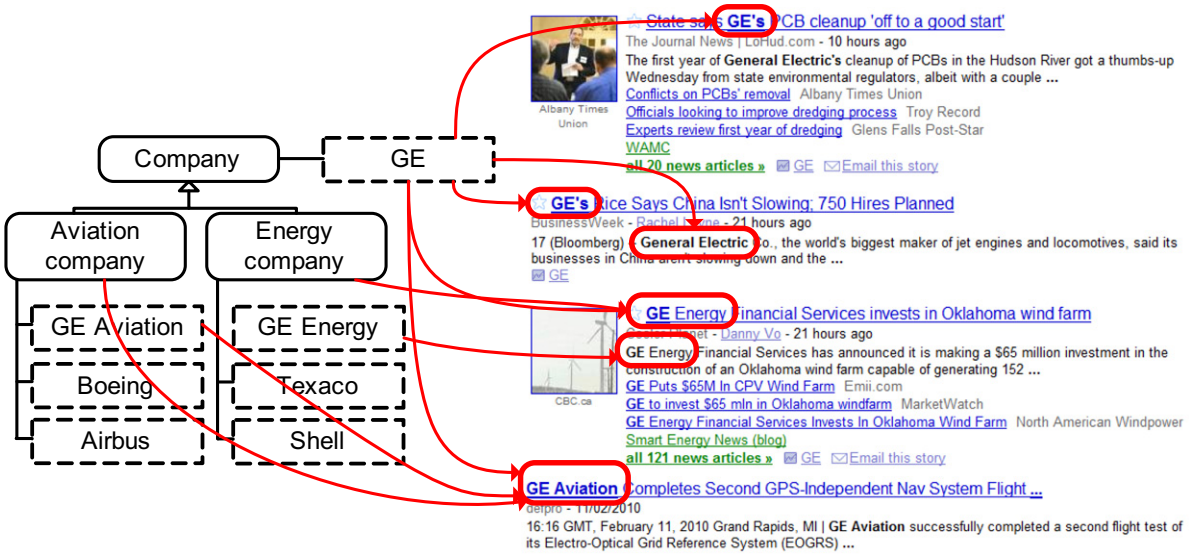

Fig. 5. Example of knowledge entities tied with news.

Company: Aviation and Energy. The results are then sorted using the semantic similarity functions described above.

4. Use case scenario: Spanish stock market

The ontology-based search engine described in the previous section has been customized to deal with financial news about the Spanish stock market domain. The ontology population module has been configured to obtain the data contained in the Madrid stock exchange market (http://www.bolsamadrid.es/). The information available from this Web data source is particularly focused on the IBEX35 index (see Fig. 6).

The IBEX 35 (an acronym of Iberia Index) is the benchmark stock market index of the Bolsa de Madrid, Spain's principal stock exchange. Initiated in 1992, the index is administered and calculated by Sociedad de Bolsas, a subsidiary of Bolsas y Mercados Españoles (BME), the company which runs Spain's securities markets (including the Bolsa de Madrid). The IBEX 35 index is made up of the 35 securities quoted on the Joint Stock Exchange System of the four Spanish Stock Exchanges (Madrid, Barcelona, Bilbao and Valencia), which were most liquid during the control period pursuant to the terms of this regulation.

The way the ontology-based semantic search engine operates in this context can be divided into two main phases: annotation and search. During the first stage, which is launched periodically, the system obtains financial news (in Spanish) from the Internet and annotates them with knowledge entities from the financial ontologies. Once this process is completed, all the gathered news

ACCIONES DEL IBEX 35

Martes, 23 de Febrero de 2010 (17:36)

Índice	Anterior	Último	Dif. (%)	Máximo	Mínimo
▼ IBEX-35	10.570,50	10.312,90	-2,44	10.638,60	10.267,50

Nombre	Últ.	Dif. (%)	Máx.	Mín.	Volumen	Efectivo (miles)	Fecha	Hora
▼ ABENGOA	19,3600	-2,22	19,9500	19,2250	302.829	5.909,25	23/02/2010	Cierre
▼ ABERTIS SE.A	13,7800	-1,50	14,1050	13,6800	1.424.816	19.737,69	23/02/2010	Cierre
▼ ACCIONA	80,8100	-1,75	82,9000	80,5500	279.592	22.795,90	23/02/2010	Cierre
▼ ACERINOX	12,7900	-0,78	13,1550	12,7300	1.481.468	19.152,06	23/02/2010	Cierre
▼ ACS	33,3400	-0,91	34,0300	33,0250	532.922	17.813,27	23/02/2010	Cierre
▼ ARCELORMITTA	28,2500	-2,45	29,4200	28,0600	581.401	16.693,59	23/02/2010	Cierre
▼ BA.POPULAR	4,8350	-2,22	4,9990	4,8000	9.371.665	45.725,45	23/02/2010	Cierre
▼ BA.SABADELL	3,5530	-1,85	3,6580	3,5300	4.301.628	15.453,38	23/02/2010	Cierre
▼ BA.SANTANDER	9,4850	-4,10	9,9690	9,4040	59.776.610	574.282,27	23/02/2010	Cierre
▼ BANESTO	7,4200	-2,37	7,6330	7,4000	404.997	3.023,19	23/02/2010	Cierre

Fig. 6. Example of IBEX-35 price.

Fig. 7. Screenshot of the semantic search engine.

is attached with the relevant semantic content that is referred to in the news.

The second phase comprises users issuing queries to the search engine. The natural language queries are analyzed and the inner meaning extracted. Then, the information regarding a particular query is matched against the financial ontology in order to determine the knowledge entities that are concerned with the query. Finally, the news that had been previously attached to those knowledge entities is returned to the users as a result of their queries (see Fig. 7).

5. Conclusions and future work

Web search engines work by storing information about many web pages, which they gather from the Web itself. These pages are retrieved by a Web crawler, i.e. an automated Web browser which follows every link it sees. The contents of each page are then analyzed to determine how they should be indexed using, for example, words extracted from the titles, headings, or other fields called meta tags. Data about web pages are stored in an index database for use in later queries. When a user enters a query into a search engine (typically by using key words), the engine examines its index and provides a listing of best-matching web pages according to its criteria. With the aim of search engines is to facilitate as much as possible the process of discovering the intended information. Since their invention in the nineties, search engines have been a complete success, as demonstrated by the presence of several search engines among the most visited websites according to the traffic rank elaborated by Alexa (http://www.alexa.com/).

However, the approach followed by traditional search engines suffers from practical limitations. First, the ever-increasing size of the Web prevents users, who are presented a large list of links as

a result of their query, from finding what they are looking for in a timely fashion. This problem can be partially overcome by developing topic-specific search engines. Examples of these are the solutions provided by Google to search for financial information with Google Finance, news with Google News, videos with Google Videos, books with Google Books, etc. A further limitation of traditional search engines is that simple keyword-based queries often return vast amount of irrelevant information which produces the low precision and recall problem. Besides, keyword-based search engines present other serious problems such as language ambiguity (synonymy and polysemy). All of this leads us to the need for better search strategies. Semantic Search Engines are a relatively recent phenomenon, supported by the Semantic Web trend.

Semantic Search is a process used to improve online searching by using data from semantic networks to disambiguate queries and web text in order to generate more relevant results. It is the logic-based underpinnings of the Semantic Web which enables the intelligent retrieval of data and helps in dealing with semantic heterogeneity. In this work, we propose a semantic search engine specially suited for dealing with financial news. We have focused on the financial domain due to the increasingly pressing demand for financial data management. In addition, the stock exchange market is a knowledge intensive domain, there are many businesses and companies involved, and therefore the economic impact could be significant.

The semantic-based search engine for financial domain proposed here provides a complete set of new features such as (i) multilingual or internationalization support, (ii) synonym inclusion, (iii) a semantic friendly search interface, and (iv) filtering results through their semantic meaning. All of them are important; yet the possibility of filtering the results should be highlighted because this will reduce the size of the results set, which improves search quality and reduces search costs for companies. However, obtaining the set of semantic annotations for the financial news is critical to improving the performance of traditional searches based on match searching.

We are currently working on upgrading this system and converting it into a recommendation system in which the users could set their preferences and the system would return only relevant news. Furthermore, this approach can be easily applied to different domains, and we are analyzing the possibility of offering our search services to the cloud (SaaS).

Acknowledgements

This work has been partially supported by European Union through the EUREKA Initiative (∑!4989), and the Spanish Ministry for Industry, Tourism and Commerce through projects SITIO (TSI-0204000-2009-148), GO2 (TSI-020400-2009-127), SONAR (FIT-340000-2007-212) and SONAR II (TSI-020100-2009-263).

References

Alonso, S. L., Bas, J. L., Bellido, S., Contreras, J., Benjamins, R., & Gomez, J. M. (2005). DIP WP10: Case study eBanking D 10.7 financial ontology. *Data, Information, and Process Integration with Semantic Web Services*. Available from http://www.14d10ip.semanticweb.org/documents/D10-7-Stock-Market-Ontology.pdf Retrieved 11.05.2009.

Aussenac-Gilles, N., Despres, S., & Szulman, S. (2008). The TERMINAE method and platform for ontology engineering from texts. In Paul Buitelaar & Philipp Cimiano (Eds.), *Dans: Bridging the gap between text and knowledge – Selected contributions to ontology learning and population from text* (pp. 199–223). IOS Press.

Bizer, C., Heath, T., Ayers, D., & Raimond, Y. (2007). Interlinking open data on the web. In *Demonstrations Track, 4th European Semantic Web Conference, Innsbruck, Austria*. Available from http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkingOpenData.pdf. (Last accessed July 2011).

Castells, P., Foncillas, B., Lara, R., Rico, M., & Alonso, J. L. (2004). Semantic web technologies for economic and financial information management. *The Semantic Web: Research and Applications*, 473–487.

Dietze, H., & Shoreder, M. (2009). GoWeb: A semantic search engine for the life science web. *BMC Bioinformatics, 10*(10).

Esmaili, K., & Abolhassani, H. (2006). A categorization scheme for semantic web search engines. In *Proceedings of ACS/IEEE International conference on computer systems and applications* (pp. 171–178).

Fox, M. S., Barbuceanu, M., Gruninger, M., & Lin, J. (1997). An organization ontology for enterprise modelling. *Simulating organizations: Computational models of institutions and groups*. Menlo Park CA: AAAI/MIT Press, pp. 131–152.

García-Manotas, I., Lupiani, E., García-Sánchez, F., & Valencia-García, R. (2010). Populating knowledge based decision support systems. *International Journal of Decision Support System Technology (IJDSST), 2*, 1–20.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199–220.

Guo, Q., & Zhang, M. (2009). Semantic information integration and question answering based on pervasive agent ontology. *Expert Systems with Applications, 36*, 10068–10077.

Langville, A. N., & Meyer, C. D. (2006). *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press.

Lei, Y., Uren, V. S., & Motta, E. (2006). Semsearch: A search engine for the semantic web. In *Proc. 5th international conference on knowledge engineering and knowledge management managing knowledge in a world of networks. Lect. notes in comp. sci.* (pp. 238–245). Podebrady, Czech Republic: Springer.

Maedche, A., & Staab, S. (2004). Ontology learning. In S. Staab & R. Studer (Eds.), *Handbook on ontologies, international handbooks on information systems* (pp. 173–190). Springer.

Maedche, A., & Staab, S. (2002). Measuring similarity between ontologies. In *EKAW '02: Proceedings of the 13th international conference on knowledge engineering and knowledge management. Ontologies and the Semantic Web* (pp. 251–263). London, UK: Springer-Verlag.

Pan, A., Raposo, J., Alvarez, M., Carneiro, V., & Bellas, F. (2007). Automatically maintaining navigation sequences for querying semi-structured web sources. *Data & Knowledge Engineering, 63*(3), 795–810.

Partridge, C., & Stefanova, M. (2001). A synthesis of state of the art enterprise ontologies – Work in progress. Lessons Learned. The BORO Program. Available from: http://www.boroprogram.org/boro_program/pdfs_trap/OES-01.pdf Retrieved 11.05.2009.

Pinsker, R., & Li, S. (2008). Costs and benefits of XBRL adoption: early evidence. *Communications of the ACM, 51*(3), 47–50.

Ruiz-Martínez, J. M., Castellanos-Nieves, D., Valencia-García, R., Fernández-Breis, J. T., García-Sánchez, F., Vivancos-Vicente, P. J., et al. (2008a). Accessing touristic knowledge bases through a natural language interface. In *Proc. PKAM 2008, Hanoi Vietnam* (pp. 147–160).

Ruiz-Martínez, J. M., Miñarro-Giménez, J. A., Guillén-Cárceles, L., Castellanos-Nieves, D., Valencia-García, R., García-Sánchez, F., et al. (2008b). Populating ontologies in the etourism domain. In *Proc. international conference on web intelligence and intelligent agent technology, IEEE/WIC/ACM, Sydney, Australia* (Vol. 3, pp. 316–319).

Seong-Bae, P., Sang-Soo, K., Sewook, O., Zooyl, Z., Hojin, L., & Seong Rae, P. (2008). Target concept selection by property overlap in ontology population. *International Journal of Computer Science, 3*(1), 14–18, 2008.

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & Knowledge Engineering, 25*(1–2), 161–197.

Sugibuchi, T., & Tanaka, Y. (2005). Interactive web-wrapper construction for extracting relational information from web documents. In: *WWW '05: Special interest tracks and posters of the 14th international conference on World Wide Web* (pp. 968-969). New York, NY, USA: ACM.

Tao, C., & Embley, D. W. (2009). Automatic hidden-web table interpretation, conceptualization, and semantic annotation. *Data & Knowledge Engineering, 68*(7), 683–703.

XBRL International (2009). XBRL: eXtensible Business Reporting Language. Retrieved June 19, 2009, from XBRL International Web site: http://www.xbrl.org.

Yang, S. (2009). OntoPortal: An ontology-supported portal architecture with linguistically enhanced and focused crawler technologies. *Expert Systems with Applications, 36*, 10148–10157.