ERMIS: Extracting Knowledge from Unstructured Big Data for Supporting Business Decision Making

Christos Alexakos(✉), Konstantinos Arvanitis, Andreas Papalambrou,
Thomas Amorgianiotis, George Raptis, and Nikolaos Zervos

Industrial Systems Institute, ATHENA Research and Innovation Centre,
Patras Science Park Building, Platani, 265 04 Rio, Patras, Greece
{alexakos,nzervos}@isi.gr, konstantinos.arvanitis@gmail.com,
andreas@papalambrou.gr, amorgianio@ceid.upatras.gr,
george.raptis@ymail.com

Abstract. Business managers support that decisions based on data analysis are better decisions. Nowadays, in the era of digital information, the accessible information sources are increasing rapidly, especially on the Internet. Also, the most critical information for business decisions is hidden in a large amount of unstructured data. Thus, Big Data analytics has become the cornerstone of modern Business Analytics providing insights for accurate decision making. ERMIS (Extensible pRoduct Monitoring by Indexing Social sources) system is able to aggregate unstructured and semi-structured data from different sources, process them and extracting knowledge by semantically annotating only the useful information. ERMIS Knowledge Base that is created from this process is a tool for supporting business decision making about a product.

Keywords: Big Data · Business Analytics · Ontologies · Data driven decision making · Knowledge extraction

1 Introduction

During the last decade the term Big Data has the pride of place in both academia and industry on the area of Business Analytics [1]. Big Data analytics methodologies and technologies permit the processing of large amount of data for providing accurate insights for a business [2]. Especially today, the century of digital information, businesses can collect useful data from various sources such as Internet Sites, blogs, social media and IoT infrastructure, even from their own information systems. This not only means the need for processing of large volumes of data but also the necessity for the analysts to face the fact that these data volumes are increasing in high rates [3]. Since, the industry sector supports that decisions based on data analysis are better decisions, the utilization of Big Data analytics enables managers to conclude to decisions based on evidence rather than intuition [4].

Another immense challenge of Business Big Data analytics is the extraction of useful knowledge from the millions of unstructured data existing in various sources on the

L. Iliadis and I. Maglogiannis (Eds.): AIAI 2016, IFIP AICT 475, pp. 611–622, 2016.
DOI: 10.1007/978-3-319-44944-9_54

Internet. This problem is entitled as Variety and it is one of the three major challenges in Big Data, called the three Vs of Big Data, alongside with Velocity and Volume [5]. Big data can be characterized in three types: (a) structured, (b) semi-structured and (c) unstructured. Structured data is provided in an already tagged and easily sorted format. Unstructured data is random thus it is difficult to be processed. Semi-structured data has separated data elements but they do not conform to fixed fields [6]. When analysts undertake problems of Business Intelligence that require multi-domain and multi-source knowledge such as the processing of crowd opinions about a product or spot consumer's behavior, the information data sources are usually web sites, posts in forums or social networks, blogs, reviews and news on portals [7]. In such cases the really useful information is "hidden" in unstructured data among other non-critical information [8].

The main scope of this article is to introduce an intelligent integrated system – called ERMIS (Extensible pRoduct Monitoring by Indexing Social sources) - which allows the digital mediation to the optimum decision making, via the intelligent process of large volume of structured and semi-unstructured data derived from various Internet sources. The specialized technological aim is the development of an intelligent layer that will integrate useful information from various sources in a unified Knowledge Base. The users of this system are both the consumers as well as the decision making persons of a business. Consumers can get information about a specific knowledge on a product by setting a query in natural language. The business managers can monitor the information for a specific product or service, by studying the extracted knowledge from a large amount of sources on the Internet.

ERMIS allows users to compose a query related to a product in natural language (i.e. "I want to buy DELL X345 laptop", "The monitor in my iPhone 6 has been broken"). The system processes the query and exports the main semantics that characterized it. It tries to recognize the product, the producer and the purpose of the query (i.e. if an article refers to damage or an intentional buy). In the background, the system collects data from the Internet (social media, news portals, e-shops, etc.), it processes them, annotates them based on an integrated ontology and feeds them to a unified Knowledge Base. Knowledge Base keeps only the useful information based on the semantics of the ERMIS integrated ontology. The extraction of knowledge related to a user's query comes with the semantic inference to the axioms stored in the Knowledge Base. The extracted knowledge is presented to the user through a user-friendly web interface. ERMIS system is designed and implemented for processing text data in Greek language, a most challenging effort due to the diversity and variety of the grammar and syntax of Greek language. Nevertheless, the system can be easily adapted for the English language and for any other language.

The proposed approach is presented in detail in the rest of the paper. In Sect. 2 some significant technologies and related work in the area of Big Data analytics based decision making are presented. Section 3 describes the basic functional components of the proposed architecture and Sect. 4 presents in detail the ontological model of the ERMIS Knowledge Base. Section 5 describes the data collection, processing and semantic annotation, while Sect. 6 refers to the knowledge extraction and presentation to the users. Finally, Sect. 7 concludes the paper.

2 Big Data and Data-driven Decision Making

A lot of decisions in the industry are based on the analysis of data. This practice is called Data-driven decision making (DDD). Decision makers can choose between two practices: in the first, more traditional, the managers based their decisions on their experience and their intuition, in the second, the managers take advantage of the analysis of business-related data in order to interpret the market trends. The second one is based on DDD techniques and it is supported by various Data Analytics tools. As DDD is not an all-or-nothing practice and it can be easily combined with the practices based on manager's experience, it is gaining the confidence of industry in the last decade [9]. A study by Brynjolfsson, Hitt and Kim [10] shows that one standard deviation higher on the DDD scale is associated with a 4–6 % increase in productivity and also affects higher return on assets, return on equity, asset utilization, and market value.

Big data technologies, especially from the side of data engineering, permit analysts to process large volume of data which leads them to more accurate decisions. It is remarkable, that a study presented by Tambe [11] shows that utilization of Big Data technologies correlate with productivity growth that can reach 1–3 % higher productivity for one standard deviation higher utilization of big data. Specifically, one standard deviation higher than the average business.

In the last years a lot of Big Data Analytics tools have been proposed by the key players in data analytics. IBM big data platform [12] and SAS Big Data Insights[1] are some paradigms of platforms that provide Big Data engineering techniques to their customers for creating analysis processes and reports. Also, most of the cloud providers have services for Big Data analysis such as Microsoft's Azure HDInsight[2] and Amazon Web Services Big Data platform[3]. Regarding the academia, the proposed DDD approaches are mainly driven to solve current problems as higher accuracy in data mining and data visualization in various fields. Visual analytics for Big Data is a big challenge as they provide users a friendly way for analysing their data. The use of visual analytics is the center of many research works such as network bandwidth evaluation for security vulnerabilities detection [13] and human muscles movement and forces simulation for diagnosis purposes [14]. Also, the proposed BIG [15] is a Multi Agent System for collecting data, unstructured text processing and decision making. BIG text process in based on keyword extraction and it does not support natural language querying.

3 ERMIS Architecture

The ERMIS system targets two user models. Everyone can create a Public User account, which allows one to make queries and receive answers. Both the queries and the answers will be visible to every user and only a single pending query is permitted for each public

[1] http://www.sas.com/en_us/insights/big-data.html.
[2] https://azure.microsoft.com/en-us/services/hdinsight/.
[3] https://aws.amazon.com/big-data/.

user account. Specific users can also create Industry User accounts, which allows the Industry User to perform multiple concurrent queries and keep the query and results private and linked to their account. In addition, Industry Users have access to Monitoring queries (queries which continue to provide answers while they remain active), notifications when their queries are answered or new data become available on a Monitoring query and the ability to generate statistical reports based on their queries. In addition, Industry User accounts are provided with access to a REST service API to allow integrating the ERMIS system with their internal software.

The ERMIS system is composed of two main subsystems the Front-End and the Back-End, each with specific component modules as presented in Fig. 1.

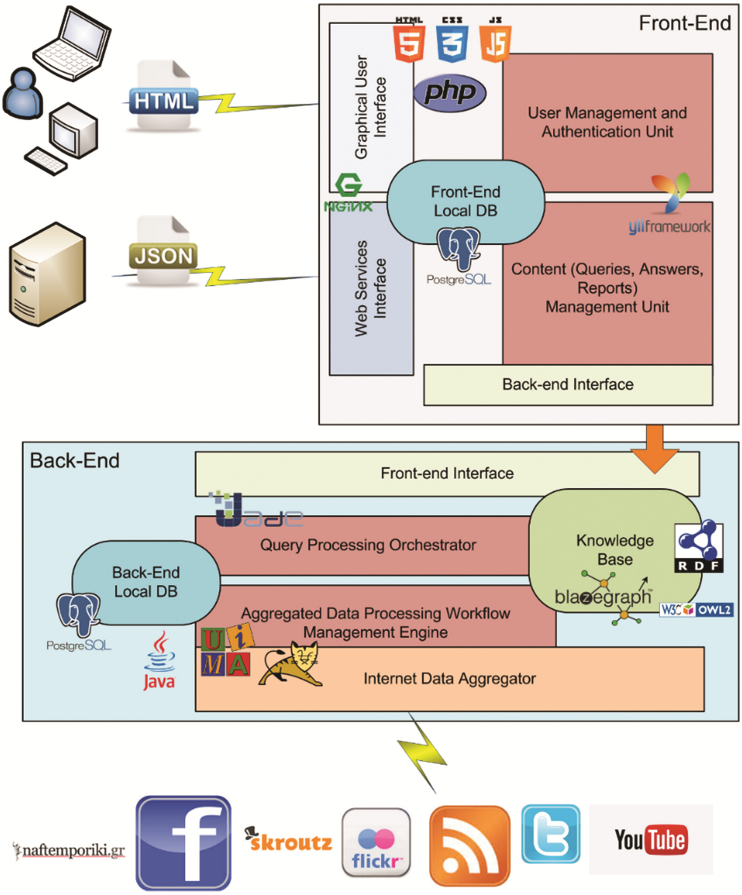


Fig. 1. ERMIS architecture

The **Front-End** components are a UI component, a user management component, a query management component and a Web Service component. The UI component is responsible for providing a web-based graphical User interface that allows interaction with the ERMIS functionality. The user management component is responsible for user authentication and authorization for the whole ERMIS system. The query management component is responsible for accepting and handling the users' queries, the system's answers to them and all the various monitoring data and reports, by scheduling tasks in the backend subsystem and handling the back-end's answers. The Web Service component provides a REST API that makes the ERMIS functionality available to third party systems.

The **Back-End** components are databases for structured and unstructured data, the Knowledge Base, a Query Processing Orchestrator and the Aggregated Data Processing Workflow Management. The databases are used to keep track of data such as reports and queries, as well as metadata for the various information sources and structured data related entities in the Knowledge base. The Knowledge Base contains the knowledge gathered through external sources, expressed in axioms composed in following RDF format and according to the OWL ontologies that are described in the next chapter. The Query Processing Orchestrator is a Multi-Agent System and it is responsible for scheduling all tasks related to query processing, such as analysis to extract relevant terms and periodic checks to update monitoring queries. It compiles the reports and answers that are then made available to the Front-End. The data processing workflow provides constant monitoring of select RSS feeds and other data sources, such as site-specific information APIs and links located through processing of the various feeds. It utilizes syntactic and lexicographic analysis of data from information sources and extracts relevant metadata and terms to store in the Knowledge base.

The Front-End is installed in a web server (NGINX[4]) and is developed on top of Yii[5] PHP framework. The provided screens to the users are empowered with HTML5, CSS and JavaScript and designed following responsive design patterns utilizing the Bootstrap[6] framework. The Back-End is developed in Java, as a set of interoperating software modules. To provide for future extension and alternative implementations, Apache UIMA (Unstructured Information Management Architecture)[7] is used to create and manage the analysis engines used for natural language processing for both queries and unstructured data downloaded from other information sources. Source specific modules are used to take advantage of site-specific APIs (such as Skroutz[8], Twitter[9] and YouTube[10]) that provide structured or semi-structured data. It uses PostgreSQL to store structured data and BlazeGraph[11] to provide the Knowledge Base functionality. Blazegraph is a high performance graph database [16] platform that supports RDF/SPARQL with scalable solutions including embedded, High Availability, scale-out, and GPU-acceleration.

[4] http://nginx.org/.
[5] http://www.yiiframework.com/.
[6] http://getbootstrap.com/.
[7] https://uima.apache.org/.
[8] http://developer.skroutz.gr/.
[9] https://dev.twitter.com/rest/public.
[10] https://developers.google.com/youtube/.
[11] https://www.blazegraph.com/.

4 ERMIS Knowledge Base

ERMIS's knowledge management is accompanied by a Knowledge Base where collected information related to products is stored in RDF graph format. The structure of the graph along with the rules defining the interrelationships are based on an integrated ontology model represented in OWL. This model defines the basic system ontology and a group of secondary ontologies and taxonomies focusing on expanding the main ontology in order to define axioms and entities which add and interrelate information gathered from multiple internet sources.

4.1 ERMIS Ontology

ERMIS ontology is the main ontology of the Knowledge Base. It describes the concepts of the Product, Document (information source) and Query. The concepts of ERMIS ontology are depicted in Fig. 2.

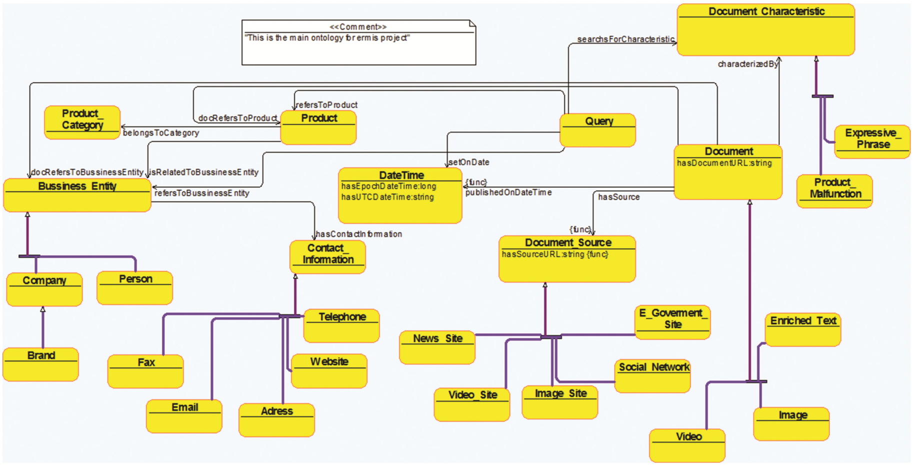


Fig. 2. ERMIS ontology

The primary classes of the ontology are:

- *Product:* This class defines a product, and is interrelated with other classes of the system. The relationship *isRelatedToBussinessEntity* is used to correlate products with business entities, such as suppliers and manufacturers, and the relationship *belongsToCategory* is used to classify the products into categories.
- *Bussiness_Entity:* This class defines all the business entities which are interrelated to a product, through its subclasses such as Person and Company. The contact information of each business entity, such as address, telephone and email address, is defined by the class *Contact_Information* and the relationship *hasContactInformation*.

- *Document:* This class defines the entities providing information related to products, collected from multiple internet sources. Each instance could be an enriched text or a multimedia object, such as image or video, as it is defined in the corresponding subclasses. Multimedia could be part of a document and this axiom is represented in the *includedInDocument* relationship. Documents' characteristics include expressive phrases, article polarity and referred product malfunction. Hence, each document refers to a product and/or is interrelated to a business entity. It is collected from different types of internet sources, such as news websites, image or video galleries, e-government services and social networks.
- *Query:* This class represents the queries set by the ERMIS users. When a query is set, a search for related documents starts based on characteristics, referred product and referred business entity.
- *DateTime:* This class defines the date and time of an action and is used when a document is published or when a query is set.

4.2 Integrated Ontology

The integrated ontology ErmisIO is a generic ontology created to combine the aforementioned ontologies of the system. All the required ontologies were imported in ErmisIO and different namespaces were used to distinguish them. The additional ontologies except from ERMIS ontology are:

- Dublin Core from Protégé for the description digital objects[12].
- GPT Ontology derived from Google Product Taxonomy used at Google Merchant[13].
- SIOC (Semantically-Interlinked Online Communities)[14] ontology for social media content
- Twitter Engineering Ontology for describing twitter content [17].

In particular, for the ontology DUBLIN CORE the annotation properties *dc:title, dc:creator, dc:format, dc:language* and *dc:description* were added. To support the ontology Google Product Taxonomy the class *ermisio:GooglePT_Entity*, a subclass of the *ermisonto:Product_Category* class, was created. The *ermisio:GooglePT_Entity* class is the superclass of all the classes on the first level of the Google Product Taxonomy ontology, and hence the taxonomy is provided as *ermisonto:Product_Category*. For the ontology SIOC the class *ermisio:SIOC_Entity* was created, which is the superclass of all the classes on the first level of the SIOC. The *sioc:Post* class is the subclass of *ermisonto:Product*, which classifies a post as document as it is defined in the ERMIS ontology. Similarly, for the Twitter Engineering ontology, the *twtronbto:Tweet* class is the subclass of *ermisonto:Product*. In the same way, other domain ontologies that describe structured data from an internet source can be integrated in ERMIS Integrated Ontology.

[12] http://dublincore.org/.

[13] https://support.google.com/merchants/answer/1705911?hl=en.

[14] http://rdfs.org/sioc/spec/.

4.3 WordNet Ontology

WordNet [18] is a lexical database of English words, which groups the words into synonyms sets, called synsets. Each synset represents a distinct lexical meaning, provides short definitions and is connected to various lexical and semantical relationships. It was created in 1985 by G.A. Miller [19] who was inspired by artificial intelligence experiments trying to understand the human semantic memory. Its main purpose was to provide a combination of dictionary and thesaurus features to support the automatic text analysis in interfacial intelligence applications.

Balkanet[15] expanded the number of European languages developed by EuroWordNet[16]. The Greek WordNet was established by the Databases Lab (DBLab) of the University of Patras along with the participation of the University of Athens [20]. The biggest ambition of BalkaNet is the semantic connection of the words of each language in order to create a multilingual semantic network.

The ERMIS project requires words recognition in a gathered text and their correlation with a product malfunction or deficiency, along with the polarity of the information (positive, negative, or neutral). For these purposes, the Greek WordNet schema was expanded in order to annotate each synset with any related information. A brief ontology was created for this purpose and is presented in Fig. 3 and its primary classes are the Malfunction and Polarity classes.

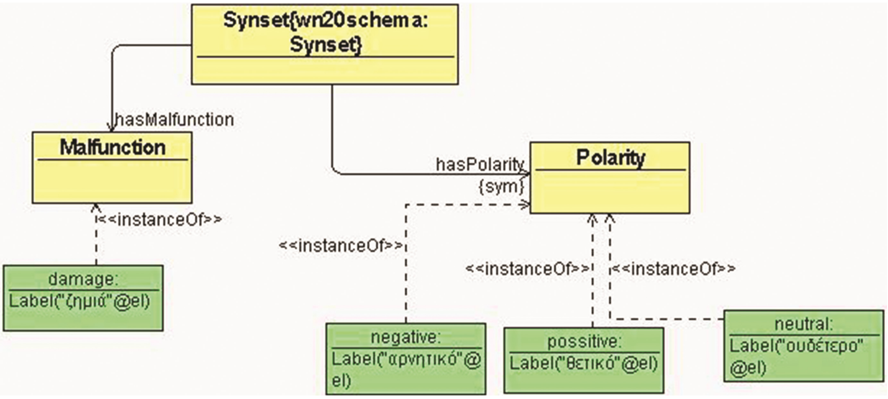


Fig. 3. WordNet extension for ERMIS project

5 Process Unstructured Data Towards to Semantic Annotation

Unstructured information found on the Internet typically exists in texts that can be formatted (containing heading, bold text etc.) or not. One of the goals of the ERMIS system is to process these texts and correlate them semantically to the ErmisIO ontology.

[15] http://www.dblab.upatras.gr/balkanet.

[16] http://www.illc.uva.nl/EuroWordNet/.

In order for this process to take place, every document fetched from the Internet (web page, forums posts etc.) is processed as depicted in Fig. 4. In more detail, the stages are as follows:

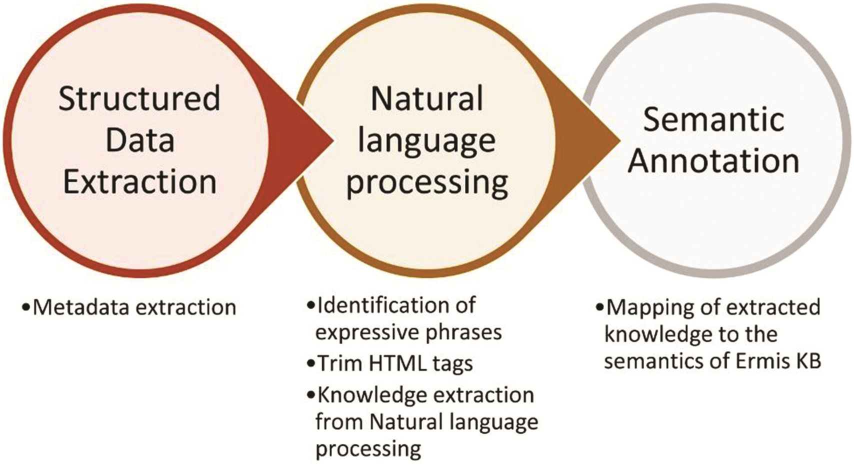

Structured Data Extraction
Natural language processing
Semantic Annotation
•Metadata extraction
•Identification of expressive phrases
•Trim HTML tags
•Knowledge extraction from Natural language processing
•Mapping of extracted knowledge to the semantics of Ermis KB

Fig. 4. Process unstructured data towards to semantic annotation

Structured data extraction: Each retrieved document is typically accompanied by metadata such as dates, authors, titles etc. For example, Twitter tweets can provide time and date, user and HTML links. HTML documents typically contain metadata information in the HEAD section. These metadata can be directly correlated with the Dublin-Core annotation properties.

Natural language processing (NLP): NLP involves the processing and understanding of natural language (in our case, in written form). In order to perform the NLP, the Apache UIMA framework is used. UIMA uses Analysis Engines to annotate documents of unstructured information and provide metadata. The following steps are needed from the retrieval of a document to the semantical analysis of the text.

- *HTML cleanup:* This involves the removal of HTML tags leaving only plain text. However, important information that could assist in the semantic analysis (such as bold, italics etc.) is kept in the form of UIMA annotations.
- *Grammatical and syntactic processing:* This involves the recognition of grammar and syntax tokens as well as meaningful tokens in the written text. First, basic grammatical processing in the form of sentence and word annotation takes place. Then words are recognized regarding their Parts of Speech and are also stemmed in order to correlate all possible grammatical forms to a specific semantic token. Finally, Named entities are also recognized.

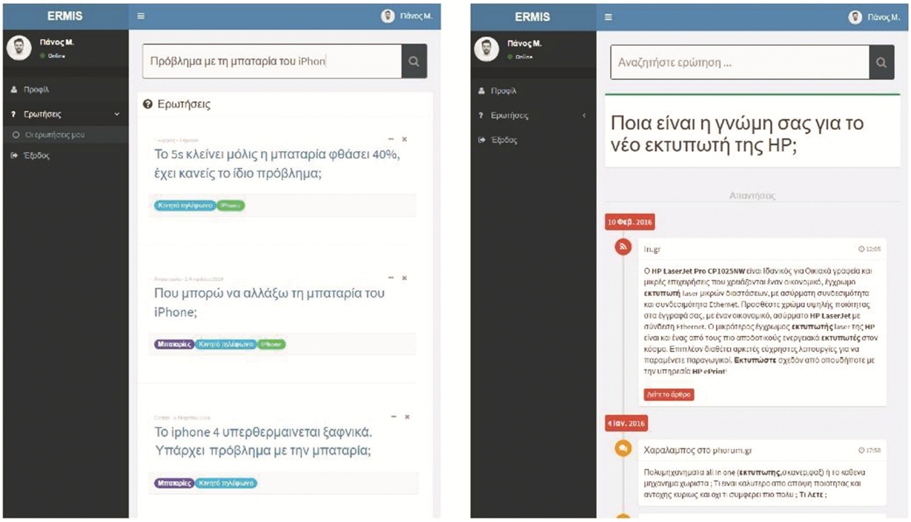
Semantic Annotation: This stage involves the mapping of the above annotated properties to the semantics of ERMIS knowledge base. The mapping involves the following specific stages.

- *Metadata mapping*. Extracted metadata from the structured data, for example dates, are converted to RDF triplets and mapped to the ontology.
- *HTML tag mapping*. Extracted HTML tags that provide useful information are converted to RDF triplets. For example, bold (< b>) tags are mapped to expressive phrases.
- *NLP results mapping*. NLP annotations are mapped to entities. For example, some nouns can be correlated to Product Categories, named entities can be correlated to companies or places, other nouns or verbs can be semantically correlated to Wordnet synsets such as a product malfunction or a positive/negative opinion on a product.

6 Knowledge Extraction

ERMIS system supports users to compose their query in natural language. When a query is set to the system, the latter is processing it in the same way as described in the previous chapter for free text information sources.

From the query the annotator is trying to extract the product, producer, product category, keywords, the type of the question (question for damage or general question) and the damage of the product if it exists. With this information the appropriate SPARQL queries are composed for getting information from the Knowledge Base. The results are evaluated in a rank system of 1 to 3 grade (match, high match and great match) according to their closeness to query's initial information. Finally, the results are presented to the user in a timeline from the most recent to the oldest one. Each answer is marked with



Πρόβλημα με τη μπαταρία του iPhon

Ερωτήσεις

Το 5s κλείνει μόλις η μπαταρία φθάσει 40%, έχει κανείς το ίδιο πρόβλημα;

Κινητό τηλέφωνο iPhone

Που μπορώ να αλλάξω τη μπαταρία του iPhone;

Μπαταρίες Κινητό τηλέφωνο iPhone

Το iphone 4 υπερθερμαινεται ξαφνικά. Υπάρχει πρόβλημα με την μπαταρία;

Μπαταρίες Κινητό τηλέφωνο

Αναζητήστε ερώτηση ...

Ποια είναι η γνώμη σας για το νέο εκτυπωτή της HP;

In.gr 12:05

Ο **HP LaserJet Pro CP1025NW** είναι Ιδανικός για Οικιακά γραφεία και μικρές επιχειρήσεις που χρειάζονται έναν οικονομικό, έγχρωμο **εκτυπωτή** laser μικρών διαστάσεων, με ασύρματη συνδεσιμότητα και συνδεσιμότητα Ethernet. Προσθέστε χρώμα υψηλής ποιότητας στα έγγραφά σας, με έναν οικονομικό, ασύρματο **HP LaserJet** με σύνδεση Ethernet. Ο μικρότερος έγχρωμος **εκτυπωτής** laser της **HP** είναι και ένας από τους πιο αποδοτικούς ενεργειακά **εκτυπωτές** στον κόσμο. Επιπλέον διαθέτει αρκετές εύχρηστες λειτουργίες για να παραμένετε παραγωγικοί. **Εκτυπώστε** σχεδόν από οπουδήποτε με την υπηρεσία **HP ePrint**!

Χαραλαμπος στο phorum.gr 17:58

Πολυμηχανηματα all in one (**εκτυπωτης**,σκανερ,φαξ) ή το καθενα μηχανημα χωριστα ; Τι ειναι καλυτερο απο αποψη ποιοτητας και αντοχης κυριως και οχι τι συμφερει πιο πολυ ; **Τι λετε ;**

Fig. 5. Presentation of extracted knowledge to the user

different icon and color according to its ranking. Figure 5 depicts two screenshots from the ERMIS system, one with the list of user's queries and one with the timeline of the answers for a specific query.

7 Conclusion

The adaptation of Data-driven decision making to modern businesses has been proved to provide significant added value to business growth. When the problem comes to the analysis of large volume of Big Data the challenges range from data engineering tasks to data explanation and knowledge extraction. ERMIS (Extensible pRoduct Monitoring by Indexing Social sources) system offers an integrated solution for collecting, processing, semantically annotating structured and unstructured data accumulated from various sources over the Internet. The main scope of ERMIS is to elaborate the natural language processing techniques for both providing users a simple way to set their queries and extracting valuable knowledge which is hidden in unstructured data sources such as web sites, posts in forums or social networks, blogs, reviews and news on portals. ERMIS system introduces an intelligent layer that integrates useful information from various sources in a unified Knowledge Base. The users of ERMIS system, which are both the consumers and the industry decision makers, are able to compose a query related to a product in natural language and view the inferred knowledge in a user-friendly web interface.

Although ERMIS supports state-of-art algorithms for Natural Language Processing, it can be extended with methods for extracting more knowledge from text processing such as sentiment analysis. Also, during the evaluation, we spotted a lot of mismatching values caused by the complexity of Greek language. In a future version of ERMIS system, we are expecting to resolve such issues. Furthermore, ERMIS is going to be extended to more information sources with both new data processing annotators and domain definition ontologies. Finally, the adaptation of new languages such as English and French is in the ERMIS' future roadmap.

References

1. Chen, H., Chiang, R.H.L., Storey, V.C.: Business intelligence and analytics: from big data to big impact. MIS Q. **36**(4), 1165–1188 (2012)
2. Wixom, B., et al.: The current state of business intelligence in academia: the arrival of big data. Commun. Assoc. Inf. Syst. **34**(1), 1 (2014)
3. Gantz, J., Reinsel, D.: The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east. In: IDC iView: IDC Analyze the Future, pp. 1–16 (2012)
4. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J., Barton, D.: Big data. Manage. Revolution Harvard Bus. Rev. **90**(10), 61–67 (2012)
5. Singh, S., Singh, N.: Big data analytics. In: International Conference on Communication, Information and Computing Technology, Mumbai, India, October 2011. IEEE (2012)
6. Sagiroglu, S., Sinanc, D., Big data: a review. In: 2013 International Conference on Collaboration Technologies and Systems (CTS), pp. 42–47. IEEE (2013)

7. Kambatla, K., Kollias, G., Kumar, V., Grama, A.: Trends in big data analytics. J. Parallel Distrib. Comput. **74**(7), 2561–2573 (2014)
8. boyd, d., Crawford, K.: Six provocations for big data (21 September 2011). In: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011. SSRN: http://dx.doi.org/10.2139/ssrn.1926431
9. Provost, F., Fawcett, T.: Data science and its relationship to big data and data-driven decision making. Big Data **1**(1), 51–59 (2013)
10. Brynjolfsson, E., Hitt, L.M., Kim, H.H.: Strength in numbers: how does data-driven decision making affect firm performance? Working paper, SSRN working paper (2011). SSRN: http://ssrn.com/abstract=1819486
11. Tambe, P.: Big data know-how and business value. Working paper. NYU Stern School of Business, NY, New York (2012)
12. Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J., Corrigan, D.: Harness The Power of Big Data. The IBM Big Data Platform. McGraw Hill Professional, New York (2012)
13. Kalamaras, I., Papadopoulos, S., Drosou, A., Tzovaras, D.: MoVA: a visual analytics tool providing insight in the big mobile network data. In: Chbeir, R., Manolopoulos, Y., Maglogiannis, I., Alhajj, R. (eds.) AIAI 2015. IFIP AICT, vol. 458, pp. 383–396. Springer, Heidelberg (2015). doi:10.1007/978-3-319-23868-5_27
14. Stanev, D., Moschonas, P., Votis, K., Tzovaras, D., Moustakas, K.: Simulation and visual analysis of neuromusculoskeletal models and data. In: Chbeir, R., Manolopoulos, Y., Maglogiannis, I., Alhajj, R. (eds.) AIAI 2015. IFIP AICT, vol. 458, pp. 411–420. Springer, Heidelberg (2015). doi:10.1007/978-3-319-23868-5_29
15. Lesser, V., Horling, B., Klassner, F., Raja, A., Wagner, T., Zhang, S.X.Q.: BIG: an agent for resource-bounded information gathering and decision making. Artif. Intell. **118**(1–2), 197–244 (2000)
16. Sikos, L.F.: Graph databases. In: Mastering Structured Data on the Semantic Web, pp. 145–172. Apress, Berkeley (2015)
17. Allen, T.: Twitter ontology, Report, National Center for Ontological Research (2013). http://ncor.buffalo.edu/2013/IE500/Reports/Travis-Allen-Twitter-Ontology.docx. Accessed
18. Miller, G.A.: WordNet: a lexical database for English. Commun. ACM **38**(11), 39–41 (1995)
19. Miller, G.A.: Wordnet: a dictionary browser' in information in data. In: Proceedings of the First Conference of the UW Centre for the New Oxford Dictionary. University of Waterloo, Waterloo, Canada (1985)
20. Stamou, S., Nenadic, G., Christodoulakis, D.: Exploring Bal-kanet shared ontology for multilingual conceptual indexing. In: LREC, European Language Resources Association (2004)