Semantic integration of web data for international investment decision support

Boyan Simeonov[1], Vladimir Alexiev[1], Dimitrios Liparas[2], Marti Puigbo[3], Stefanos Vrochidis[2], Ioannis Kompatsiaris[2]

[1] Ontotext Corp, Bulgaria
{boyan.simeonov, vladimir.alexiev}@ontotext.com
[1] Information Technologies Institute, Centre for Research and Technology Hellas, Thermi-Thessaloniki, Greece
{dliparas, stefanos, ikom}@iti.gr
[3] PIMEC, Barcelona, Spain
mpuigbo@pimec.org

Abstract. Given the current economic situation and the financial crisis in many European countries, Small and Medium Enterprises (SMEs) have found internationalisation and exportation of their products as the main way out of this crisis. In this paper we provide a decision support system that semantically aggregates information from many heterogeneous web resources and provides guidance to SMEs for their potential investments. The main contributions of this paper are the introduction of SME internationalisation indicators that can be considered for such decisions, as well as the novel decision support system for SME internationalisation based on inference over semantically integrated data from heterogeneous web resources. They system is evaluated by SME experts in realistic scenarios in the section of dairy products.

Keywords. decision support, indicators, heterogeneous web resources, SME internationalisation, semantic integration

1 Introduction

Given the current economic situation and the financial crisis in many European countries, SMEs have found internationalisation and exportation of their products as the main way out of this crisis. To this end, the SMEs need to find relevant information that will facilitate this process such as: i) spending habits of consumers in potential markets, ii) economic fundamentals of the countries (indicators micro and macro), iii) geographic and entry barriers (legislation, certifications, etc), iv) consumer behaviour, v) domestic and foreign competition, vi) distributors of its product to export in the selected markets, vii) contact information of potential customers. In order to find this information the SMEs have to access foreign trade offices in each country (e.g. Chambers of Commerce), dedicated databases (e.g. Market access database , Eurostat, etc.), as well as the web by using general purpose search engines (e.g. Google). In these resources, the companies expect to find the competence, potential clients and in general all the information that is required to take a decision for exporting products to the right country. However, this method is time consuming and complicated, because information is distributed and heterogeneous and there is no existing platform that provides access to all the necessary information. The additional problem faced by many SMEs is the language barriers, since this information is usually provided only in the language of the host country. To deal with this problem, we need technologies that provide unified access to multi-lingual and multicultural economic material across borders in order to guide the international investments of SMEs.

In this paper we present a decision support system that semantically aggregates information from many heterogeneous web resources and provides guidance to SMEs for their potential investments. The main contributions of this paper are the introduction of SME internationalisation indicators that can be considered for such decisions, as well as the novel decision support framework for SME internationalisation based on inference over semantically integrated data from heterogeneous web resources. To the best of our knowledge this is the first attempt to develop a decision support system for SME internationalisation.

The rest of this paper is organised as follows: Section 2 provides some theoretical background, as well as an overview of the related work. Section 3 describes the SME internationalisation indicators that are utilised by the decision support framework presented in this paper. The different components of the proposed framework are described in detail in Section 4. In Section 5, the experimental results from the application of the framework to a number of datasets collected from several resources are presented. Finally, some concluding remarks are provided in Section 6.

2 Related work

Decision support systems (DSS) can be broadly defined as computer-based applications that support people and organisations in their decision-making processes. Research on this very important scientific field has spanned 50 years and many different kinds of systems have been presented. According to [1], DSS can be divided into the following main categories:

- Model-driven DSS: These include computerised systems that employ accounting and financial models, representational models, and/or optimization models to assist in decision-making [2]. One representative system in this category is ILOG JRules [3].
- Data-driven DSS: These systems aim at accessing and processing large amounts of data. Simple file systems accessed by query and retrieval tools provide the most elementary level of functionality in this category [4]. A nice example of data-driven DSS is the Geographical Information System (GIS), which can be used to visually represent geographically dependent data using maps. It should be noted that the framework proposed in this paper belongs to the data-driven DSS category.
- Document-driven DSS: Multimedia document collections serve as the backbone of the decision-making process in document-driven DSS. Document

analysis and information retrieval (IR) systems are simple examples from this category [5].

- Communication-driven DSS: These systems aim at supporting groups of people working on a given task, by focusing on the interaction and collaboration aspects of decision-making. At its basic form, a system of this type can be a simple threaded e-mail and in its complex form, it can be an interactive video or a web communication application.
- Knowledge-driven DSS: They actually recommend or suggest actions to the users, rather than just retrieve information relevant to a certain decision, i.e., these systems try to perform some part of the actual decision-making for the user through special-purpose problem-solving capabilities [5].

A detailed literature review with respect to the research conducted on the use of semantic web technologies in the DSS context can be found in [5]. A large number of semantic web-related studies have focused on the medical and healthcare domains. For instance, [6] introduce a semantic web framework that efficiently models the knowledge within clinical practice guidelines (CPG), in order to develop a knowledge-centric clinical decision support system (CDSS). In addition, [7] explore the use of Web Ontology Language (OWL) reasoning services, in order to execute CPG in CDSS. In [8], a generic architecture for the semantic enhancement of CDSS, which also considers the reutilisation of knowledge embedded in a CDSS, is proposed. Another prominent research domain is e-commerce. In this context, [9] introduce a Semantic Web Constraint Language (SWCL) based on OWL and utilise it, in order to implement a shopping agent in the Semantic Web environment. Furthermore, [10] design and develop a shopbot that can help customers compare products located in e-stores, using different languages. In order to achieve this, they propose a semi-automatic method of constructing multilingual ontologies, as well as a semantic searching mechanism based on concept similarity. In another approach [11] the authors present a system that provides high quality environmental information for personalized decision support based on reasoning.

In addition to the aforementioned works, there have been some DSS-related studies that deal specifically with financial management and investment decision-making. More specifically, [12] employ the Object Oriented Bayesian Knowledge Base (OOBKB) design to develop a real-time DSS that supports managing of investment portfolios. In another work, [13] presents a hybrid intelligent system that consists of a DSS based on portfolio management rules, as well as a fuzzy inference system. Finally, a detailed analysis of tools implemented in DSS to support individuals in their financial management and investment decisions is provided in [14].

This paper, inspired by the ontology-based decision support systems such as [9] and [11] presents a knowledge-driven DSS for SME internationalisation based on semantic integration of heterogeneous internet data.

3 SME internationalisation indicators

In most cases, the main issue for SMEs to internationalise is to assess the different countries that could be potentially interesting for exporting their products. The selection of the correct country for the international investment depends on a number of indicators, which allow a comparison and therefore helps the decision-making process. In order to build the decision support tool, we need to combine the indicators in a sophisticated manner. To do that, we have to initially conduct a screening and establish a categorisation of the indicators so that we can prioritise and weight them according to their relevance. The grouping considered captures a framework of macro-environmental criteria, which is considered in the strategic management of SMEs when assessing opportunities or threats. The study for the definition of indicators has taken place in the context of MULTISENSOR project[1] with the support of the SME internationalisation department of PIMEC[2].

To analyse, select and organise the indicators and its grouping, we considered the PEST analysis together with other indicators more focused on the product. SMEs export managers with a vast experience on internationalisation assessment participated in the elaboration of the methodology of the decision support. As we are dealing with targeting foreign markets, the decision support tool needs to consider external factors. In this sense, the PEST analysis refers to the combination of Political, Economic, Social and Technological factors which can affect the business. Within this framework, contrasted indicators from reliable sources – e.g. Eurostat[3], World Bank[4] - were selected so that the final result is robust.

For a more complete and personalised feature, SME export managers agreed that there was the need for another group of indicators that directly related to the small company that is looking for a market to export. For this purpose, we included UN Comtrade data which measures trade for every product and set of products. Thus, the decision support tool incorporates the selected product data so that the outcome is not only the result of the countries' general factors but also, and decisively, of the concrete product commercialisation.

Altogether, we obtain a combination between product specific and personalised data with country factors. Indeed, every indicator and group of indicators have a different importance when making a decision. Hence, the decision support tool integrates a differentiated weight for every indicator and category according to SME export managers' criteria. In more detail, we grouped the indicators into 4 categories:

- Product: This is the key category in our system, as it is directly connected with the product the SME is producing or offering. Here, we include UN Comtrade data which gives precise information of the export/import flows between countries worldwide, segmented by product in the Harmonised System code. We take into account trade between the targeted country and the rest of the world, as well as the trade with the country of origin of the SME running the decision support tool. Furthermore, we incorporate a variable that measures the trade flows per capita. Last, distance between countries is included; a value that is given different weight depending on the product, as the type of good conditions the importance of a fast delivery. In all, the Product category cap-

[1] http://www.multisensorproject.eu/
[2] http://www.pimec.org/
[3] http://eurostat.linked-statistics.org
[4] http://worldbank.270a.info/

tures the balance of trade of the specific products and brings very relevant information to compare the economic and consumption specificities of the countries in relation to the SMEs' commercialisation.

- Economy: This category combines macroeconomic data that gives an input of the countries' recent economic performance – e.g. GDP growth, GDP per capita - together with general balance of payments data. Also, the Easiness Doing Business ranking from the World Bank plays a major role, as it describes how difficult it is to do business in every country, including variables such as how complex it is to start a business, to what extent contracts are enforced and how easy is to trade across borders. Hence, we obtain a sum of the health of the countries' economies, its overall trade flows and its business and legal culture.
- Politics: Membership of economic areas and trade agreements constitute the basis of this category. In this sense, the main variables are being or not a member of the European Union, the European Economic Area or the European Free Trade Association, which affects decisively in commercial trade. Additionally, legal certainty aspects are also included with indexes that reflect political stability, levels of corruption and government effectiveness.
- Social: This category has the lower weight and embraces societal aspects that tell us the market dimension and consumption possibilities of the countries with selected indicators such as their total population, their level of education or their unemployment rate. Furthermore, we include an economic perception index that gives us the sentiment that consumers have towards their economy, which can affect their preferences when it comes to their consumption.

In Table 1 we present the most important SME internationalization indicators categorised and weighted.

4 Decision Support Framework

The Decision Support System is composed of these main components: Indicator information mining from the web, semantic integration of this data in a semantic repository (database) and the decision support mechanism. Then the guidance is presented through a user friendly interface.

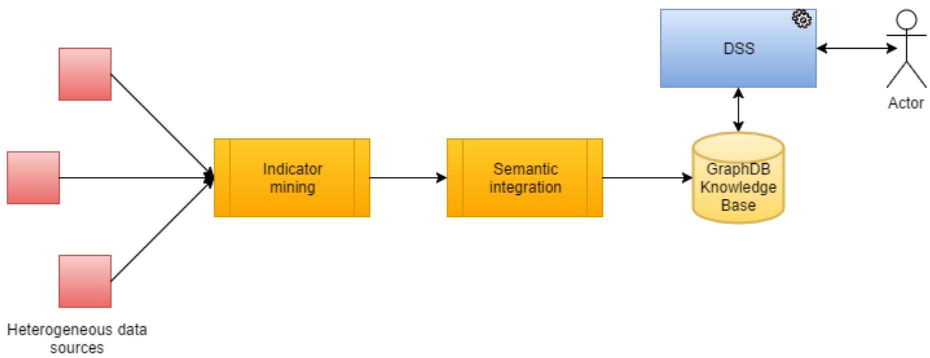

Fig. 1**.** Data flow (architecture) of the decision support system.

4.1 Web retrieval and mining of indicator data

In order to extract information on the SME internationalisation indicators we use dedicated APIs from specific websites such as Eurostat and WorldBank and UN COMTRADE (see section 5.1 for details). Then we convert this information to RDF (section 5.2) and loaded them in Ontotext GraphDB[5], a semantic repository that can store and query semantic data. We selected the W3C CUBE ontology [16], which is the most appropriate way of representing statistical data in RDF format. We also use elements from SDMX [17], which provides terms for some common dimensions such as population, GDP, etc.

Some sources are already available in the required format [18], e.g. data from Eurostat and World Bank. For other sources, e.g. UN COMTRADE[6], we had to use a specific API. We used the free API that has the following limits: 100 requests per hour per IP address, and maximum of 50k records per request (each record represents one trade flow between two partners for one product group). This provides plenty of data (potentially up to 120 million records per day), but we had to make sure that none of our requests exceeded the limit of 50k. We did this through judicious selection of dimensions and downloading strategies (e.g. how many Year series and Product codes to retrieve at a time). To collect this data we developed a program that queries the COMTRADE API repeatedly by varying the parameters (see section 5.1.3), downloads the data and converts it to the required format.

In addition, we used the Google distance API to extract information regarding city distances in various transportation modes (e.g. air vs surface). We used the Google distance API, which is free for 2.5k requests. In order to stream-line processing, we pre-fetched the distances between the capitals of all source and destination countries, and converted to RDF using a custom program.

With respect to other indicators such as Government type, Political instability, Corruption percent index, Human development index, etc., we downloaded them manually from various web sites and converted them using a mix of automatic and manual approaches. They are provided in different tabular formats, such as web pages, csv, tsv. After manual cleanup if required, we converted these data to RDF.

4.2 Semantic data integration

Data integration from disparate data sources is often required for online analytical processing (OLAP) and DSS analyses. In recent years, semantic data integration [15] has emerged as the most promising integration approach, because of the simple and uniform data model that it uses (RDF). RDF is a graph data mode, in which data is broken into atomic facts called triples. URLs are used to identify every block of data, and every property (relation or attribute). This allows data sharing on a global scale.

Reusing property names and values defined by others (in our case, SDMX and Eurostat) is one of the benefits of the semantic web. It both reduces the time required to

[5] http://ontotext.com/products/graphdb/

[6] http://comtrade.un.org/

create data models, and the chance of data reuse by others. After converting the data to RDF and loading it to Ontotext GraphDB, we calculate some derived data using SPARQL UPDATE [20]. Then, for each pair of observations IMP/EXP with matching dimensions, we calculate two derived observations: TOT (total trade) and BAL (trade balance). We record them in URLs that mirror the original URLs, where the Indicator part is replaced appropriately.

Since we deal with many indicators in different areas, each indicator has its own value range and direction of growth (for some increase is desirable, for others decrease is desirable). To perform meaningful calculations over this heterogeneous data we need to normalize data to the same range, using this formula:

$$z_i = x_i - min(x) / max(x) - min(x)$$

4.3 Query-Based Decision Support

After statistical data is converted to RDF and loaded into the Knowledge base, it can be used for decision support. The user selects his country of origin and the desired commodity code (product group). We use the following mechanisms:

- Weighted sum of indicators (see Table 1 for the weights, which are empirically selected).
- Simple inference rules, e.g. "If the commodity code indicates a perishable product then use air cargo distance; if it's a heavy product then use surface cargo distance; else compare both".
- Simple decision tree elements.

We implement these mechanisms using parameterized SPARQL templates to retrieve the appropriate statistical data from the cube. This backend processing allows us to implement the following features:

- Rank target countries (and select the most appropriate one) for the
- Compare two countries across all indicators. E.g. we can compare Germany and Austria across GDP, GDP per capita, corruption level, human development index, population purchasing power, etc.
- Display various charts to illustrate the time dimension (we use the Google Charts API).

Using these functionalities, a manager or entrepreneur can obtain solid information supporting to support the decision making process.

5 Results and evaluation

5.1 Dataset descriptions

The following data were retrieved and integrated for this experiment: a) Eurostat dataset, including statistical indicators for Europe**;** b) World Bank dataset, which included World Bank Indicators; c) UN COMTRADE, which offers comprehensive data on cross-country trade volumes for various types of goods, collected since 1962. Other indicators obtained from specific sources such as: i) Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically motivated violence, including terrorism. (Worldwide Governance Indicators - The World Bank Group); ii) Corruption perception index (Transparency International). In the table below you can see the main selected indicators, divided by categories. Each indicator is assigned an individual weight (w1). The indicator categories are also assigned with weight w2.

Table 1. SME internationalisation indicators and weights

Group	Indicator	w1	w2
Economy	GDP growth	6	6
	GDP per capita PPP	5	
	Exports in % of GDP	3	
	Imports in % of GDP	3	
	Inward FDI stocks in % of GDP	3	
	Export Import ratio	3	
	Harmonised Index of Consumer Prices (Inflation)	3	
	Easiness of Doing Business	10	
	Average days to Export	6	
	Average days to Import	6	
Social	Total population	8	3
	% Tertiary education	3	
	Unemployment rate	5	
	Economic Sentiment Indicator	3	
	Human Development Index	3	
	% Internet at Home	3	
Politics	EU Member	10	5
	EEA Member	10	
	EFTA Member	10*	
	Political Stability Index	3	
	Corruption Perception Index	3	
	Government Effectiveness Index	3	
Product	Export (partner: World)	5	10
	Import (partner: World)	10	
	Export (partner: country of origin)	5	
	Import (partner: country of origin)	8	
	Product balance (Export - Import / Partner: World)	8	
	Product balance (Export - Import / Partner: country of origin)	6	
	Imports per capita (Import divided by Population / Partner: World)	8	
	Imports per capita (Import divided by Population / Partner: Country of origin)	6	
	Distance	10	

5.2 Experimental setup

In order to evaluate the proposed system we have conducted experiments at PIMEC in Barcelona [21]. The evaluation involved a focus group of five (5) export managers, as well as one-on-one interviews with three (3) SME export responsible and one (1) export manager. This focus group was used to test the individual functionalities of the system. First we provided an initial explanation of the status of the system and described the functionalities that were available for testing. The participants were asked to evaluate a table of indicators where the user could compare two countries and visualise their differences. Specifically the task was the following: each user was supposed to be the CEO of an SME selling dairy products (e.g. cheese and yogurt), who wanted to decide which country is the most convenient for exporting products. In fig 2 we illustrate the user interface of the decision support system in which different indicators for Slovenia and Czech Republic are compared[7].

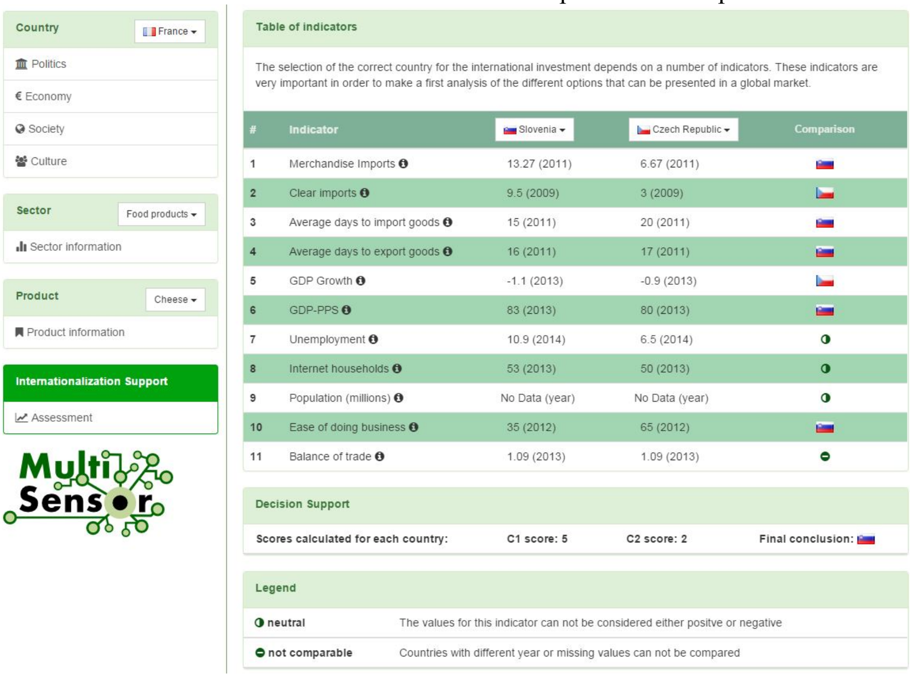

Fig. 2. Decision support interface

5.3 Results

The results of the performed evaluation are considered in general satisfactory. On the task, where the users were asked to evaluate the table of indicators and the decision support functionality the table was valued positively as a way to more easily

---

[7] A demo of the DSS is available at: http://grinder1.multisensorproject.eu/uc3/

compare the situation in two countries. Its relevance was also valued positively. Larger number of indicators and more concrete ones were mentioned as a way to better capture the advantages or disadvantages of exporting to one country or another.

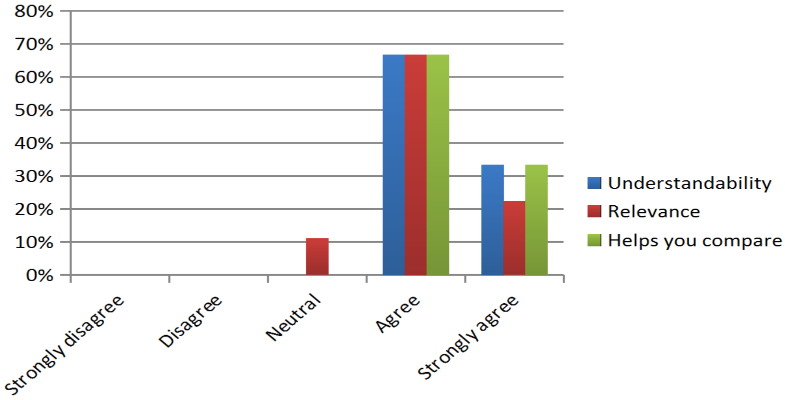

Fig. 3. Evaluation of indicators table and decision support functionality

After evaluating concrete aspects of the different tasks, the evaluators were required to value the overall efficiency of the decision support platform. The results are visualised on Figure 5 and as we can see the opinions were highly positive. All the participants felt that the system was easy to use and, also, that it allowed them to save time compared to alternative ways of looking for similar information.

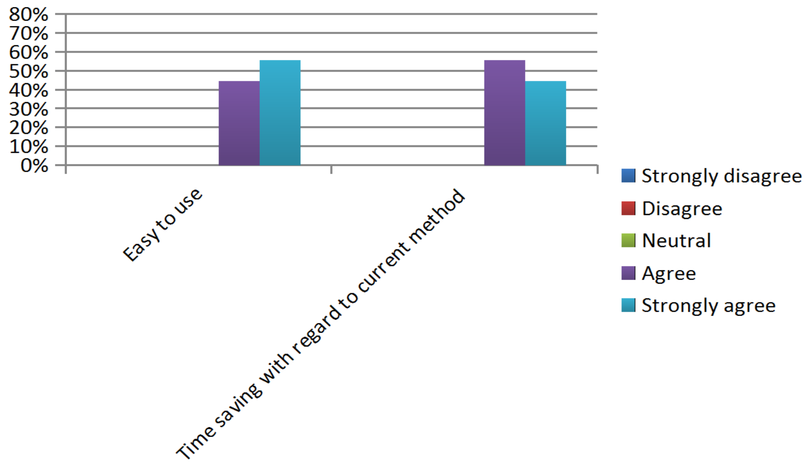

Fig. 4. Efficiency evaluation

Regarding satisfaction, the results are overall positive but there are some differences that are worth pointing out. All the participants felt in control when using the

DSS and thought that it was intuitive and easy to use. Participants appreciated the user interface and navigability of the platform. A vast majority considered the use of MULTISENSOR DSS as a satisfying experience and a way to be more productive. Most of them would also recommend the system to colleagues.

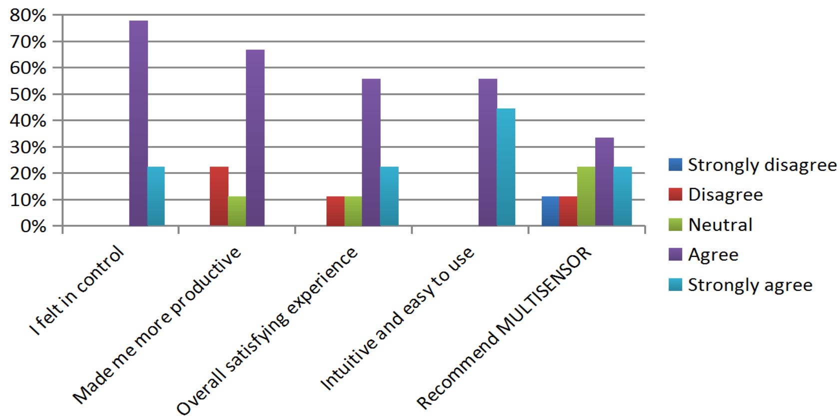

Fig. 5. Satisfaction

6 Summary and conclusions

In this work we define SME internationalisation indicators and we provide a decision support tool for SME internationalisation, which builds upon semantic integration of information from heterogeneous web resources. This application could support SMEs in order to have guidance in deciding in which country they could expert. It provides a comparative view of the countries in question and shows insights based on the SME internationalisation indicators. The evaluation with professionals working on SME internationalisation shows the potential of this tool in the market.

Future work includes the crawling of more indicators, more extended evaluation, as well as employment of additional techniques for providing guidance such as decision trees and fuzzy reasoning.

Acknowledgments. This work was supported by FP7-610411 MULTISENSOR.

References

1. Power, D. J. (2002). Decision Support Systems: Concepts and Resources for Managers. Studies in Informatics and Control, 11(4), 349-350.
2. Power, D. J., & Sharda, R. (2007). Model-driven decision support systems: Concepts and research directions. Decision Support Systems, 43(3), 1044-1061.
3. Power, D. J. (2006). What are examples of decision support systems in global enterprises?. DSS News, 7(7).

4. Power, D. J. (2008). Understanding data-driven decision support systems. Information Systems Management, 25(2), 149-154.
5. Blomqvist, E. (2014). The use of Semantic Web technologies for decision support–a survey. Semantic Web, 5(3), 177-201.
6. Hussain, S., Abidi, S. R., & Abidi, S. S. R. (2007). Semantic web framework for knowledge-centric clinical decision support systems. In Artificial intelligence in medicine (pp. 451-455). Springer Berlin Heidelberg.
7. Jafarpour, B., Abidi, S. R., & Abidi, S. S. R. (2011). Exploiting OWL reasoning services to execute ontologically-modeled clinical practice guidelines. In Artificial Intelligence in Medicine (pp. 307-311). Springer Berlin Heidelberg.
8. Sanchez, E., Toro, C., Carrasco, E., Bueno, G., Parra, C., Bonachela, P., & Guijarro, F. (2011). An architecture for the semantic enhancement of clinical decision support systems. In Knowlege-Based and Intelligent Information and Engineering Systems (pp. 611-620). Springer Berlin Heidelberg.
9. Kim, H. J., Kim, W., & Lee, M. (2009). Semantic Web Constraint Language and its application to an intelligent shopping agent. Decision Support Systems, 46(4), 882-894.
10. Huang, S. L., & Tsai, Y. H. (2011). Designing a cross-language comparison-shopping agent. Decision Support Systems, 50(2), 428-438.
11. Wanner, L., Rospocher, M., Vrochidis, S., Johansson, L., Bouayad-Aghae, N., Casamayor, G., Karppinen, A., Kompatsiaris, I., Millee, S., Moumtzidou, A., Serafini, (2015). L. Ontology-centered environmental information delivery for personalized decision support, Expert Systems With Applications, Volume 42, Issue 12, 15 July 2015, Pages 5032–5046.
12. Tseng, C. C., & Gmytrasiewicz, P. J. (2002, January). Real time decision support system for portfolio management. In System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on (pp. 1348-1356). IEEE.
13. Casanova, I. J. (2012). Portfolio investment decision support system based on a fuzzy inference system. In Computational intelligence (pp. 183-196). Springer Berlin Heidelberg.
14. Weber, B. W. (2008). Financial DSS: Systems for supporting investment decisions. In Handbook on Decision Support Systems 2 (pp. 419-442). Springer Berlin Heidelberg.
15. Alexiev, V.; Breu, M.; de Bruijn, J.; Fensel, D.; Lara, R.; and Lausen, H. Information Integration with Ontologies: Experiences from an Industrial Showcase**.** John Wiley and Sons, February 2005. ISBN: 978-0-470-01048-8
16. Cyganiak, R. and Reynolds, D. The RDF Data Cube Vocabulary. W3C Recommendation 16 January 2014. http://www.w3.org/TR/vocab-data-cube/
17. Statistical Data and Metadata eXchange. Official site for the SDMX community. https://sdmx.org/ . Accessed Apr 2016
18. Capadisli, S., Auer, S. and Riedl, R. Towards Linked Statistical Data Analysis. Proceedings of the 1st International Workshop on Semantic Statistics, Sydney, Australia, October 11th, 2013. CEUR Volume 1549, urn:nbn:de:0074-1549-5. http://csarven.ca/linked-statistical-data-analysis
19. Beckett, D., Berners-Lee, T., Prud'hommeaux, E. and Carothers, D. RDF 1.1 Turtle: Terse RDF Triple Language. W3C Recommendation 25 February 2014. https://www.w3.org/TR/turtle/
20. Gearon, P., Passant, A. and Polleres, A. SPARQL 1.1 Update, W3C Recommendation 21 March 2013. https://www.w3.org/TR/sparql11-update/
21. Heise, N., Wagner, T., Eckhoff, M., Eckhoff, M., Vrochidis, S., Peleja, F. Second Prototype Evaluation Report, December 2015. http://www.multisensorproject.eu/wp-content/uploads/2013/09/D8.4_2ndPrototypeEvaluationReport_v1.0.pdf