

Ontology supported search engine and knowledge organisation, prototyped for international niche market information

von Wolf Engelbach, Reinhard Höhn, Frauke Weichhardt

This paper describes an ontology supported software prototype that combines the advantages of existing Internet search engines with modern text analysis functionalities and an intelligent storage system for documents and knowledge items. The ontology assists the user in query definition and structures the storage of documents as well as knowledge items. The system is implemented and tested for the business case of SMEs that want to internationalise. It can easily be transferred to other domains just by changing the ontology.

Introduction

Currently it is hard to find documents on the Internet that really fit to the user's information demand, although they might exist. This paper presents a software prototype that combines the advantages of Internet search engines with modern text analysis functionalities and an intelligent ontology based storage system for documents and knowledge items.

There are three core challenges regarding the integration of search engines and knowledge management:

- First, the definition of the crucial questions for a given topic.
- Second, the intelligent integration and analysis of structured and unstructured information from the Internet.
- Third, the intuitive and complex structure of storage for detected

documents and gathered knowledge items.

All three challenges are addressed by the use of one and the same ontology in the whole software application: This greatly supports small and medium-sized enterprises (SMEs) in working on these selected results and sharing such information with colleagues and partners. This consistent and expandable concept structure

- assists in the definition of queries by topic and keyword suggestion, including all instances of an concept, and all synonyms of an instance
- is used to label and retrieve the results manually, and to classify them automatically as soon as enough results are labelled manually, and
- allows saving and retrieving of manually extracted knowledge items.

The software prototype is implemented and tested for the business case of SMEs that want to internationalise and expand their business to other countries: they need to retrieve and interpret information about their specific industry niche in a particular geographical area. Especially for such tasks, that are seldom conducted, such a predefined ontology structure helps not to forget any important issue and to store document results in a proven way. In the case of weakly structured information, as for individual industrial niches, the aspect of knowledge storage support the user extract relevant content himself.

The software was developed in an EU-sponsored CRAFT project "Analysis Of Marketing Information For Small And Medium-Sized Enterprises (AMI-SME)", and is currently named after the project. It is based on the requirements for search definition and evaluation of SMEs during their internationalisation activities:

- Differentiate companies such as partners, competitors and customers
- Identify products and its features on the market
- Allow repeated execution of searches

- Support textual analysis of the results (e.g. abstracting, classification, ...)

Business case and technical architecture

Increasing competition and globalisation trends are challenging companies to expand target markets for their products and services into foreign countries. The process of internationalisation necessitates many decisions. Adequate information about target markets is required to support decision making and ensure successful implementation of the internationalisation strategy. Among the relevant factors are market potentials, competitors, the legal situation, the sales organisation, and marketing events.

The proposed software is designed to gather, analyse and process essential information that is niche specific and often not available in editorially proven commercial or public databases, but distributed over several homepages of companies, research and governmental institutions, media, or private websites. The Internet search is done by the web services of widespread search engines, especially Google, Yahoo and a9. Metadata-extraction, filtering, clustering and ontology-based labelling functions help to keep an overview in the ocean of information pieces.

Intended users of the solution are the following players:

- SME internal personnel that is responsible for internationalisation in order to get better information faster as decision support.
- External consultants (business consultants, marketing agencies etc.) that help SMEs in their internationalisation projects and get their information more efficiently.
- Marketing intermediaries (such as business associations, chambers of commerce, public development agencies) that offer their services to

SMEs and extend their portfolio to internationalisation support.

Ontology design and usage

The ontology was modelled by the project members in Sementalk and exported to OWL (Web Ontology Language). The OWL file is loaded to a database which speeds up the storage of new concepts, instances, attributes of values during the usage of the system. The ontology can again be exported to an OWL file for further improvement within Sementalk or other modelling tools like protégé.

The user interface is optimised to offer much functionality while hiding the complexity of text-analysis and ontologies for the user. The software has a clear user guidance with main pages for project search, project documents and project knowledge. The concepts of the ontology are displayed in the user interface comparable to a folder structure in form of a tree. Instances and relations are presented in two separated lists. This allows even users without ontology experience a simple orientation and a well-known navigation.

The whole system is designed to work on several independent projects in parallel. Each project has its specific ontology. If you start a new project, a project specific copy of the system ontology is created, and all changes will be stored to this. Initially, a system ontology is delivered with the software bundle, which is oriented towards internationalisation and general marketing. In addition, it is possible to define another system ontology which is related to a different domain or topic, or you can add industry or country specific ontologies.

Defining a query in a search engine is usually done by typing a word into the adequate field of the graphical user interface (GUI). Then the search engine looks for all occurrences of this word it can find in the document corpus. You can do

the same in this software and then work on the results of that search. Moreover, the user can add different ontology labels to each search result, and also add the same labels to all selected search results. This happens in a document details pop-up (see figure 1), and here also automatically generated abstracts and extracted metadata are displayed, partly only after the download of the result document to the AMI-SME repository.

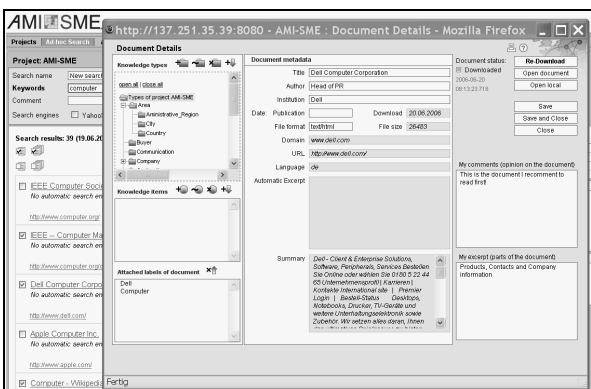


Figure 1: Document details pop-up screen

If enough results are assigned manually to a given label, this label will be used as a class for automatic classification. The user still can dismiss the suggested labelling. In the project documents screen, the user can then systematically look at all the documents that are assigned to the same label.

Moreover, also the manually extracted project knowledge is stored directly in the project ontology (see figure 2). For each instance, the values for the attributes can be entered, and new (sub-) concepts or instances can easily be defined, and relations can be assigned between the instances. This knowledge editing happens in the project knowledge view, which has the same ontology based navigation structure as the project document view.

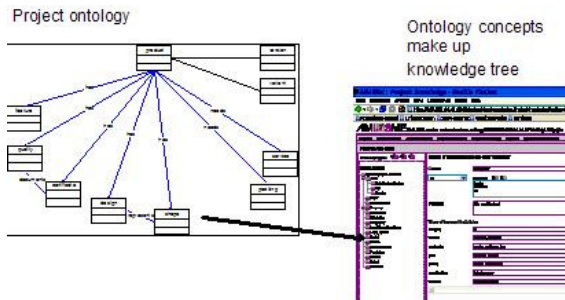


Figure 2: Ontology use in AMI-SME (project knowledge screen)

Based on all this ontology support for document and knowledge handling, AMI-SME developed an assisted search that takes ontological concepts and instances to expand the ordinary search query: The user picks a concept from the concept list and on one hand gets presented all the subclasses and instances of this concept, which he can include in his key words. On the other hand all instances and the related concepts are shown to him, so he can decide which ones he would like to include in his search term directly. Therefore search definition is performed in a more controlled and precise way, and always takes the newest items of the knowledge base into account..

Marketing Ontology

The system ontology for the AMI-SME prototype is created to support internationalisation activities of small and medium-sized companies. Their purpose is to support the strategy process for getting a better market position or to enter a new market, which heavily depends on a reliable set of information. Two well known methods of the marketing theory can help to ask the right questions and to define the ontology:

- In the first case, the company has a new product with defined features but it does not know exactly the market segment: who are the customers, how to talk with them, and what are adequate advertising measures. Here the questions are related to the optimal marketing measures for a defined product. The suitable method

is the Marketing Mix Concept by [Meffert2002].

- In the second case, the company is sceptical about the market chances, since the competition situation is not transparent: who are the competitors, what are entrance barriers, is any risk to be expected from substitutes, is there a dependency from suppliers, and who are the customers in the market. The suitable method in this case is the Competition Analysis by [Porter1980].

The marketing mix (see figure 3) is a method for definition of the status quo and also for a new status of product features. The focus of this model is the market segmentation. There are four logical steps to define the marketing components of the product (features), also called the 4 P (products - positioning - pricing - promotion): First of all the product has to be defined, with its variants, shapes and colours, features and packing, etc.. Closely related are the necessary services and qualifying certificates.

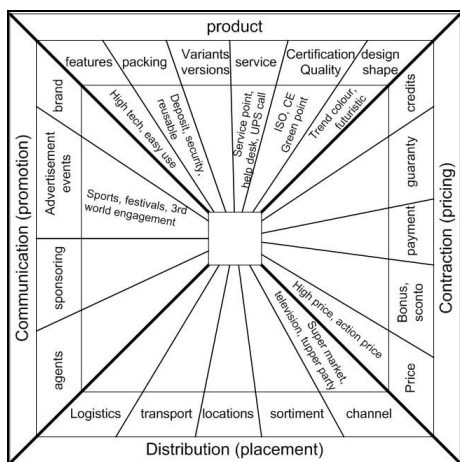


Figure 3: Marketing mix method [Meffert2002]

The next step should be the decision about the distribution with respect to the transport vehicles, locations and logistics. Then a pricing pattern with payment conditions, credit options, return conditions and bonus programs can be designed. The last step in the configuration of the marketing mix is the communication and promotion of the product to the customers, the retailers and

other target groups. The basis of the marketing mix is the action fields of the marketing. That is: product specification, the pricing features, the distribution or positioning properties and the kind of communication or promotion into the market.

The model of competition forces (see figure 4) describes the types of companies that are threatening my success. Most of the questions are related to the immediate competitors: "What are the competitors with similar products", "which channels do they use", and so on. Another important force are the buyers: "is there typical buying behaviour", "what selection procedures are known", "what typisation of consumers is useful", "has the buyer preferred locations to buy", and so on. It depends on specific project which of the other competition forces has to be analysed additionally as well. This analysis is necessary to understand the market situation and to determine marketing measures against the competition threats.

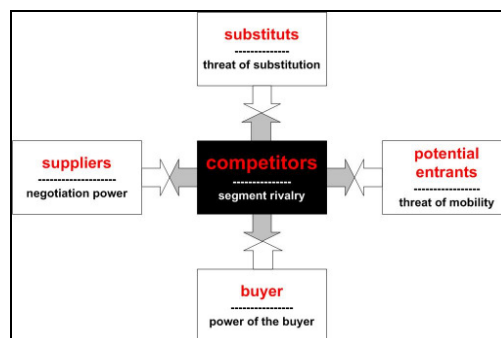


Figure 4: Competition forces [Porter1980]

Based on these two methods and some systematic approaches by [Kottler2004], the general marketing ontology is constructed (see figure 5). Its three main information fields are the competitive market situation, the product market environment, and the regional market. The picture shows only a selection of attributes to get an essential view on the concepts.

Because of the necessary relation of the competition field to the products, the part of the competition forces of the ontology consists of the concepts products of the competitors in the same market segment,

potential entrants (potential newcomers), suppliers (supply chain partners, supply power), product substitutes (that means products which can substitute features of my product), and customers (buyer). Because of the necessary relation of the marketing features of the products, the part of the marketing mix conception of the ontology consists of the concepts product (bundles, features, shape, colours, services), contraction (pricing): bonus points, rabat, product bundle, warranty, distribution (positioning) (channels, partners, locations, place, delivery), and communication (promotion, advertisement, events).

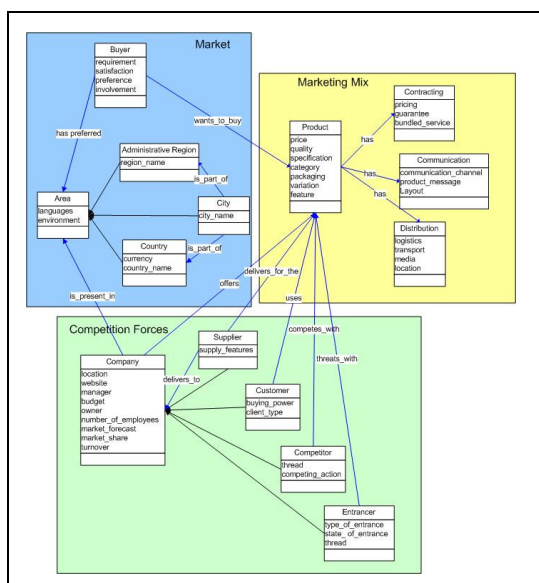


Figure 5: AMI-SME Marketing ontology

AMI-SME will be delivered with this standard marketing ontology. In order to obtain valuable search results it could be wise to augment this standard ontology with industry or company specific ontologies. This follows the approach in other domains, where generic ontologies exist, e.g. for organisational knowledge [Gualteri&Ruffolo, 2005] that can be instantiated to a specific situation, e.g. to a specific company.

This would mean to connect the standard ontology and project specific ontologies. The connection point could be, for example the concept "product" or the concept "company" in the general

ontology. Also an interim step is possible, e.g. an association delivers an industry specific ontology, and then a company specific ontology would be created, e.g. based on instances of the product and with additional attributes and relations for the concepts.

Conclusion and summary recommendations

The concept of ontology based search and storage improves the interface between search engines and knowledge management: it makes Internet search more intelligent and integrates it closely to the user's demand. In the sample application for supporting SME's internationalisation, possible end users will test the offered functionality in summer 2006. The test case will prove if ontology based functions are a good way to organise results from search engines in a complex environment.

The general idea can be transferred to other topics than internationalisation only by exchanging the ontology, e.g. product development or innovation management. For the introduction of the software in companies and associations, additional services such as technical support, as well as internationalisation and ontology consulting were designed.

Acknowledgements

The software is being developed within the EU-supported CRAFT project "Analysis Of Marketing Information For Small And Medium-Sized Enterprises (AMI-SME)" by six RTD partners and seven SME partners from five European countries. For more details see www.ami-sme.org.

References

[Gualteri&Ruffolo2005] A. Gualteri and M. Ruffolo, An Ontology-Based Framework for Representing Organizational Knowledge, In: Proceedings of I-Know 05, Graz, Austria, June 29- Juli 1, 2005, pages 71-78.

[Kotler2004] P. Kotler, Marketing Management, 2004

[Meffert2002] H. Meffert, Marketing, 2002

[Porter1980] M.E. Porter, Competitive Strategy, 1980.