In: C. Stephanidis (ed), Universal Access in HCI, Vol. 4 of the Proceedings of HCI International, Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, 2003, pp. 750-754.

Intelligent Human Language Query Processing in Mkbeem

Aarno Lehtola[1], Johannes Heinecke[2], Catherine Bounsaythip[1]

[1]VTT Information Technology
PO Box 1201, FIN-02044 VTT, Finland
Aarno.Lehtola@vtt.fi Catherine.Bounsaythip@vtt.fi

[2]France Télécom R&D, DMI/GRI/Langues Naturelles
2 avenue Pierre Marzin, F-22307 Lannion cedex, France
Johannes.Heinecke@rd.francetelecom.com

Abstract

This paper describes how ontologies are used to mediate between languages and to infer answers to user questions in the multilingual eCommerce mediation system Mkbeem. As an example, the paper discusses on how generic ontologies of colours and materials are used to infer additional facts about clothing products in order to facilitate information access. We also present how ontologies are applied in selling travelling services. The Mkbeem system prototype is in principle language independent but it has been so far tested in Finnish, French, English and Spanish.

1 Introduction

Ontologies have been widely recognised as a central solution for sharing conceptions of goods and services among parties in eCommerce (Fensel 2001, Leger & *al.* 2000, Gomez-Perez & *al.* 2001). A recent survey by IBM and Icon Medialab found that in the Nordic countries on average 35% and in Finland up to 60% of purchase attempts failed in eShops. A major cause for this bad usability was that the customers could not find the requested products. Simple string based product search facilities are not enough. "No product available" is an insufficient answer, if the selection includes comparable goods or if the user just happens to use terms that differ from the ones in the catalogue. eShops need to solve the best possible offerings matching the user requirements, like human shopkeepers would do. The required question-answering capabilities can be realised by inferring based on domain ontologies, e.g. product models, and related generic ontologies. Moreover, ontologies can be used to facilitate multilinguality.

In the Mkbeem project (Multilingual Knowledge Based European Electronic Marketplace, IST-1999-10589) ontologies serve as the central solution for providing multilinguality and intelligent question answering (Mkbeem 2000-2002). The main result of this project is a multilingual eCommerce mediation system. This Mkbeem system supports three main functionalities:

- *Multilingual Cataloguing*, which enables providers to describe in their own language the goods and services that are on sale. Textual descriptions are translated automatically. Facts about products are extracted automatically into a language neutral form that complies with the product models of the domain ontology.
- *Processing of Customer Language Information Requests*, which is based on the co-operation between *Human Language Processing* and *Ontologies* of the commerce domain, the related products and generic common sense issues. Ontologies bridge between languages and also help in implementing fuzzy information search.

- *Multilingual Trading*, which among other things applies an eCommerce ontology in carrying out contract terms adaptation for a particular shopping basket taking into account the countries of the seller and the buyer.

The Mkbeem system prototype supports currently Finnish, French, English and Spanish. The technology can be easily adapted to other languages as well. Feasibility tests have been conducted with test users since September 2002 in France and in Finland for mediating clothes, railway tickets, Finnish holiday cottage and French hotel room reservations, and car rental.

We next outline how ontologies are used to mediate between languages and to infer answers to customers' questions. After that we discuss how inference based on a colour ontology and a material ontology has been integrated in order to enrich knowledge concerning clothing. Later on we describe the use of ontologies in the travelling domain.

2 Ontologies in Sharing Product Knowledge Multilingually

The Mkbeem system uses several types of ontologies that are defined in a language neutral way. *Domain ontologies* include product models and service-ontologies defining components and properties of the mediated products. *Generic ontologies* capture general knowledge related to the products, e.g. materials. An *eCommerce ontology* defines, e.g., contracting rules. Figure 1 illustrates how ontologies are used in extracting facts from textual product descriptions and then accessing and sharing the information when a customer requests it in her native language.

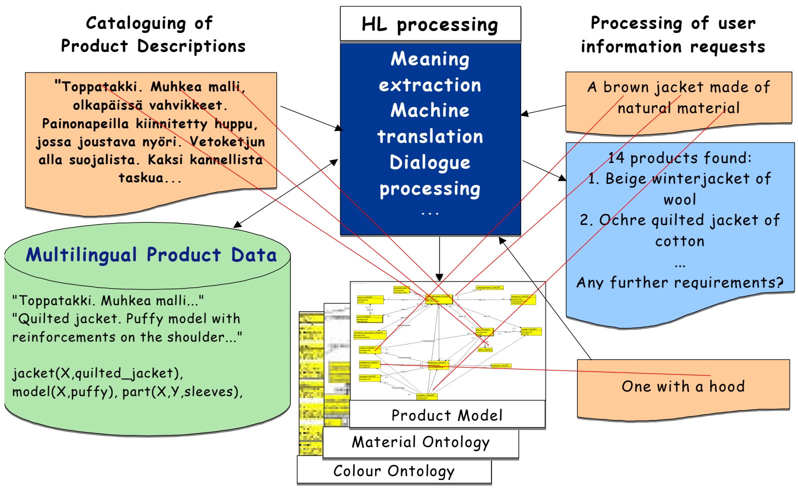

Figure 1. Using ontologies in cataloguing and accessing product information multilingually.

Both textual product articles and customers' natural language information requests are linguistically analysed. Associations to ontology concepts and found relationships are explicated using CARIN description logic language (Levy & Rousset 1998). E.g. the input query "A brown jacket made of natural material" produces for a language neutral meaning representation the *Ontological Formula* (c_colour)(X),(r_name)(X,brown),(c_product)(Y),(r_name)(Y,jacket),(c_material)(Z),(r_name)(Z,nat_mat). This formula explicates the referred ontology concepts and serves as the input for a query planning process, which includes inferring based on the ontologies. When a textual product descriptions is catalogued, the same analysis is carried out, but the corresponding Ontological Formula is used as an input for deducing ontology based property facts for the product. These facts are later on used, when customers requirements are matched to the product selection. The linguistic analysis during cataloguing is described in detail in (Lehtola & al. 2003).

3 Intelligent Question-Answering in the Clothing Domain

In the clothing domain there are ways to use common sense knowledge to help the user to find the required goods. There are basically two ways of using this knowledge. It can be used in the back-office routines while new goods are introduced for mediation. This yields automatic enriching of product data by deducing new explicit facts based on generic ontologies. Another option is to carry out the inferences during the on-line session of the customer.

3.1 Flexible Search of Products by their Colours

Colours constitute a specialised field of knowledge that has very much value to the sales of clothes, especially in a mail-ordering business with large selection. The mediation system must know more about colours and their interrelationships than just the colour name labels of the underlying product database. When the customer asks for "ochre shirt", the system should be able to find all the shirts that contain a yellow-brown shade in their colouring. Because of linguistic, cultural and perceptive differences, people have different perception of the world and naming of their perception. The colour ontology was developed to provide the e-customer a somewhat natural dialogue with the machine, e.g. flexibility regarding to linguistic and cultural diversity in the colour naming and perception.

Representing knowledge on colour names and categories is a difficult task. What is constant in every language is that colour appearance has a categorical nature (Kay 1997), e.g. there are eleven categories in English (i.e. black, white, red, green, yellow, blue, brown, purple, pink, orange and grey). The colour ontology model implemented in the project contains these eleven colours as basic colour categories. Each category is defined by its names in different languages and interval values in the CIELab colour model, specific colours are defined their names and for the moment their "exact" co-ordinates. Further on, any colour will be defined by fuzzy membership function. This will allow flexible computing of colour similarity and finding of harmonious colour sets.

When the user makes a search for a product specified by a colour name, the system will find the product with the same colour name if it exists in the database otherwise similar colours within a certain degree of similarity. When the user asks for similar colours, the system analyses the colour name and finds the category that it belongs to and retrieves products with colours that are in the same colour category. When a colour belongs to several categories, the fuzzy membership functions are of particular importance to compute similarity.

The extension of the colour question-answering system is to help users to find colour schemes.

Colour schemes could also be used to provide answers to a customer's inquiry of clothing when the customer explicitly mentions the scheme like "I want a skirt in Mediterranean colours".

3.2 Searching Clothes with their Material Properties

In the clothing area, it is an added value service for those customers who are sensitive to material properties to be able to specify in their query the type of material they want. In the last decades, new fibres have entered the market. The problem is that manufacturers and retailers have different ways of identifying their fibre products, fabrics and finish agents, e.g. the same fibre raw material may have several trade names. And even without this difference, the name of a textile material may not tell much about its properties for a common eShopper. The textile material ontology addresses this problem in order to give satisfaction for both eShops and eShoppers. On the eShop side, the ontology will enable to make junction between manufacturers data and data to provide to customers. On the customer side, the ontology will enable her/him to search for clothes that have certain functional and comfort requirements, without necessarily knowing all the fibres names and their properties. In fact, the online shoppers will be able to specify in their query for example: "*A*

T-shirt with low allergy risk" or "*a breathable sportswear*", etc. In a long run, this may even increase the customer's knowledge and awareness in their further purchases.

4 The Processing of Information Requests in the Tourism Domain

In Mkbeem, we focus on three basic services of the tourism domain, i.e. train reservation, accommodation reservation and car rental. In this area, human languages allow a wide range of expressions and the related service-ontology should contain all the necessary information related. The service-ontology is used to transform human language inputs into ontological representations. Another benefit of this is that it helps the user specify as much parameters as needed in a single request, in natural language, avoiding tiresome form-filling. This section explains how the request in natural language is processed in order to access to databases of products and services.

4.1 Linguistic Processing

To ensure that the generated language neutral Ontological Formulas will contain all relevant information given by the user, the user request is treated in several interdependent steps.
As the Mkbeem prototype is multilingual, the first step is to identify the language of the request. In the next step, the user request is analysed and a language independent semantic graph is created based on a dependency syntax tree, a set of language dependent rules comparable to that of DRS in (Kamp and Reyle 1993) and a set of language independent predicates. To ensure the ontological appropriateness of the generated semantic graph, it is checked by OntoClass (Picsel 1999-2001), which uses a linguistic domain ontology developed for this purpose. Any inappropriate semantic graph is deleted from the set of possible solutions. Finally, in order to deal with travel dates etc. (in the tourism domain) temporal expressions relative to the time of utterance (deictic elements like *now, today, in two hours, in four days, next Monday, at ten to eleven pm* or incomplete dates (*the* 12^{th} *of April, on Christmas*) are transformed into the corresponding absolute temporal expression respective to the current date (i.e. 31^{st} *May 2003*, *15:26*, *22:50*, 25^{th} *December 2003*).
The next step is a transformation of the internal semantic representation into the Ontological Formula, which is also understood by other modules. The concepts (and roles) differ considerably from the linguistic ontology due to the fact that linguistic expressions and semantic nuances are present in the semantic representation, which are not needed in the Ontological Formula. So for instance temporal/modal information (*I want to go/I would like to go/we have to be in*) must be eliminated by the transformation. Further, different lexemes expressing a move (*go/arrive/depart/travel*) need to be mapped on the concept *trip*, which is the only move-concept of the service ontology.
A typical user request is "*we would like to leave Paris for London on Monday evening"* or in French *"lundi on aimerait aller de Paris à Londres en fin de journée".* The requests inquire information on public transport from Paris to London on (next) Monday evening. The internal semantic representation for these examples and the corresponding Ontological Formula are below:

wanting(theme=x300,experiencer=x301), *goto(destination=x302,origin=x303,* *agent=x301,situation=x300)* *evening(time=x300), London(location=x302)* *Paris(location=x303), weekday~monday(date=x300)* *speaker(theme=x301), hour~18(time=x300)* *minute~0(time=x300), monthday~24(date=x300)* *month~february(date=x300), year~2003(date=x300)*	*(time)(C17),(depPlace)(V300, const_Paris)* *(arrPlace)(V300, const_London),* *(weekday)(C16, const_monday), (hour)(C17,* *18), (minute)(C17, 0), (day)(C16, 24),* *(trip)(V300), (date)(C16), (month)(C16,* *const_february), (year)(C16, 2003),* *(depDate)(V300, C16), (depTime)(V300, C17)*

5 Conclusions

This paper described how to use ontologies to enable multilinguality and to implement intelligent question answering in a multilingual eCommerce mediation system. First, we outlined how this Mkbeem system uses ontologies as a pivot to mediate between languages and to infer answers to users' questions. In the clothing area, the system can enrich its knowledge on the product properties using external ontologies, e.g., colour and material ontologies, which enables the system to find products for the user using also fuzzy criteria that are not explicit in the original product data. We also described query processing in the travelling domain where, e.g., inference about time is an important subtask. An example of how the system solves time related human language expressions was given.
The Mkbeem system prototype has been tested by the user companies which have found its functionalities to respond well to their requirements. Based on the test feedback, the system can be further extended by incorporating new capabilities such as multimodality, speech support, mobile terminal support, automatic ontology extraction etc.

Acknowledgements
The authors would like to thank all their colleagues in the Mkbeem consortium (France Telecom, SNCF, Fidal, SchlumbergerSema, UPM, NTUA, CNRS and VTT) for the excellent co-operation and the European Commission for supporting the reported work.

References

Fensel, D. (2001). *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce.* Springer Verlag, Berlin Heidelberg, 138 p.

Gomez-Perez, A., & al. (2001). *Requirement, Choice of a Knowledge Representation and Tools.* Public Report of MKBEEM project (EC IST-1999-10589), at http://www.mkbeem.com.

Heinecke, J. and Cozannet, A (2003). Ontology-Driven Information Retrieval: a proposal for multilingual user requests. In: Proceedings of Workshop on "Ontological Knowledge and Linguistic Coding" at the *25th annual meeting of the German Linguistics*, Munich, Feb. 26-28.

Kamp, H.; Reyle, U. (1993): *From Discourse to Logic. Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory* [=Studies in Linguistics and Philosophy 42.] Dordrecht: Kluwer.

Kay, P., Berlin, B., Maffi, L., Merrifield, W. (1997). Color Naming Across Languages. In: *Color Categories in Thought and Language*, C. L. Hardin and L. Maffi (eds.), Cambridge.

Leger, A., & al. (2000). Ontology domain modeling support for multi-lingual services in E-Commerce: MKBEEM. *ECAI'00 Workshop on Applications of Ontologies and Problem-Solving Methods.* Berlin, Berlin, 4 p.

Lehtola, A., & al. (2003). Multilingual Cataloguing of Product Information of Specific Domains: Case Mkbeem System. In: *Proceedings of EAMT/CLAW´03,* Dublin, May 15-17, in print.

Levy, A., Rousset, M.C. (1998). CARIN: A Representation Language Combining Horn rules and Description Logics. *Artificial Intelligence Journal*, vol. 104, September 1998.

Mkbeem (2000-2002). The Mkbeem project. URL http://www.mkbeem.com/

Picsel (1999-2001). The Picsel project. URL http://www.lri.fr/~picsel/