In: Proceedings of eChallenges-2003 - Building the Knowledge Economy: Issues, Applications, Case Studies. Part 2. Cunningham,P., Cunningham, M. & Fatelnig, P. (eds.). IOS Press (2003), pp. 947 - 954.

Multilingual and Ontological Product Cataloguing Tool – User Experiences

Aarno LEHTOLA, Tuula KÄPYLÄ, Catherine BOUNSAYTHIP, Markus TALLGREN
VTT Information Technology, POB 1201, FIN-02044 VTT, Finland
Tel: + 358 9 4561, Fax: + 358 9 4566027, Email: Aarno.Lehtola@vtt.fi

Abstract: This paper describes the multilingual cataloguing tool developed for the Mkbeem eCommerce mediation system. The very central point in the cataloguing tool is the integrated use of ontological product models and language engineering to support a set of basic cataloguing activities. The starting point for cataloguing a new product is its textual description article. After language checking of the textual description, the properties of the product are extracted from it, categories in selection are proposed for the product, and the description is translated into all supported target languages in one go. We describe the user experiences from the field trials of the cataloguing tool and discuss its possible future development.

1. Introduction

Native English speakers comprise less than 9 % of the world population. However, in March 2003 their share was still around 35 % in the Internet population and over 68 % of the WWW content was in English [2]. The proportion of native speakers of other languages is growing quickly among Internet users. The linguistic diversity is huge among them. Even within Europe there can easily be counted over 60 languages among which even the smaller ones comprise potential customer groups for international eShops. When wider social groups start to use the Internet on everyday basis, service providers cannot any more rely on foreign language skills of the users. Especially in the consumer sales native language support has been found very important. There is a remarkable need for cost effective IT solutions for enabling multilinguality in consumer Internet trading.

Ontologies combined with natural language processing can be used to facilitate the use of eCommerce services and enable their multilinguality. Ontologies provide means for sharing conceptions of goods and services among parties in eCommerce [1,3,6]. Enriched with lexical and other language models, they can facilitate multilinguality. This paper describes how these technologies are used in our multilingual cataloguing tool developed for the eCommerce mediation system Mkbeem (Multilingual Knowledge Based European Electronic Marketplace, outcome of the EU project IST-1999-10589) [5,14]. Moreover, we will present user test experiences from using our tool in cataloguing clothing products and vacation cottage rental services in Finnish, French and English.

2. Objectives

For companies dealing with thousands of product description articles in several languages, one can imagine the amount of manual and repeated work that it requires writing and translating the product descriptions. For instance, in the mail-order clothing industry, which has to follow the fashion trends and seasonal publishing of the catalogues, there is a need for an integrated cataloguing environment. This should help in producing consistent and uniform information as the whole cataloguing process is based on product models and brand-specific language terminologies that conform to the company knowledge of the domain. This knowledge can be modelled in the form of ontologies and language models.

The cataloguing process involves two types of ontologies. Domain ontologies include product models defining components and properties of the mediated products and services. Generic ontologies capture general knowledge related to the products, e.g. materials and colours. All these ontologies must be defined in a language neutral way. Related lexicons and language models need to be provided in the covered languages to support multilingual use of these ontologies.

The products and services are described in short textual articles, which may contain a photograph or other picture. It is necessary that these product descriptions are produced in the same editorial process both for the Internet and for printed catalogues. A pivot product article is maintained in one of the supported languages and it is reproduced multiple times for various media and languages. This kind of edit-once-publish-many arrangement is used both when documents are published in multiple media channels and when catalogues are translated into multiple target languages and locales.

3. Ontological Product Models and Language

The parts and properties of the products on-sale are defined in product models, or in other words domain ontologies. Figure 1 contains an example of a hierarchical product model. In the top level there is defined a very general clothing product. We call this top level model as the “metamodel”. It tells that a clothing product may contain several parts, has a model and is characterised by a colour. More details are introduced on the next level, which involves definition of properties and parts of specific types of clothes. Using subclass relation, e.g., jogging pants inherit the properties of trousers like legs and their cuffs.

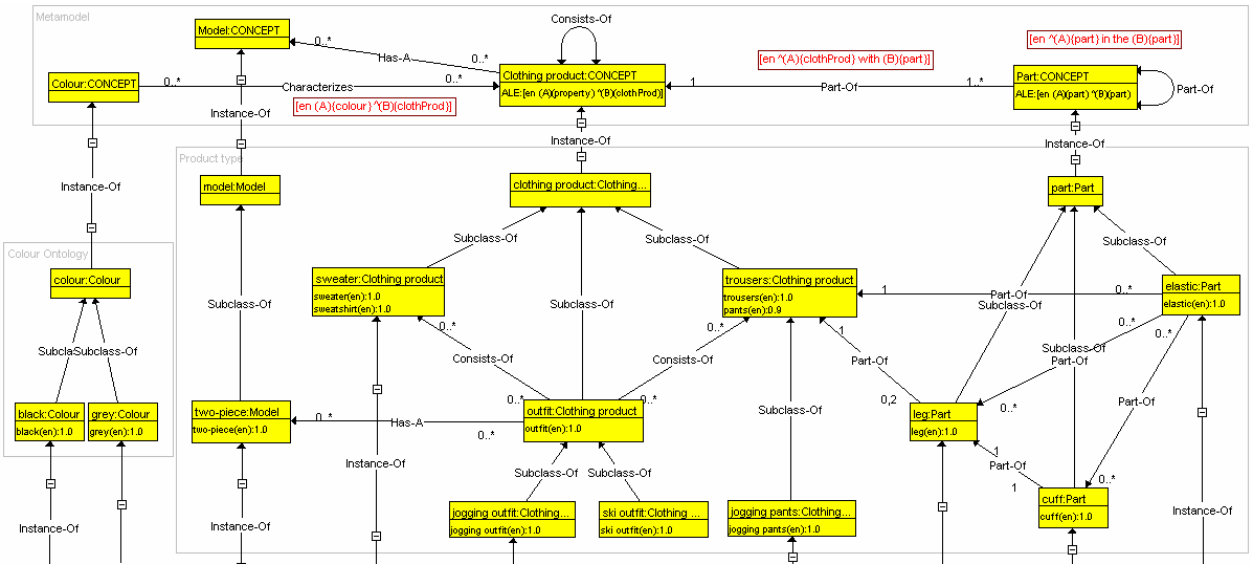

Figure 1. An example of a hierarchic product model including language models.

The models include also linguistic knowledge in the form of ALE rules (Augmented Lexical Entries) [7]. They are used to define the natural language expressions that correspond to the concepts and relations in the ontological models. Based on those one can automatically derive ontological meaning representations for input natural language expressions. These representations are called ontological formulas in the Mkbeem system. The description logic language CARIN is used for expressing the ontological models and the semantics of human language expressions [12]. The ALE rules can also be used for the generation of natural language texts on the basis of the ontologies. In other words, the use of linguistic information combined with the ontologies enables the interaction between the user and the on-line service in human language, thus making virtual shopping experience more natural.

For instance, the utterance “Two-piece jogging outfit with elastics at the leg cuffs” might be a part of the user dialog in an eShop for clothing products. Based on matching the user utterance against the terms related to the concepts of the ontology, the concepts two-piece (product model), jogging outfit (product type), elastic, leg, and cuff (product parts) are recognised.

The use of the ontological approach enables more intelligent question-answering capabilities for the system than an ordinary keyword–based search can provide [11]. Once the query semantics has been analysed with respect to the ontologies, the system can infer based on additional ontological knowledge answers to such customer questions that do not directly refer to the product fact sheets. For instance, by using a generic textile material ontology the customer service system can find products, when the customer asks for “a shirt made of natural material” or “easy-care trousers”. When the customer asks for “a khaki jacket”, the system can use a colour ontology to find the fuzzy definition of khaki colour and to infer that sand brown jackets could be appropriate, as well. As the ontologies have been defined in a language neutral way, these question-answering capabilities are available for queries in all covered languages.

4. Multilingual Cataloguing Tool

The multilingual cataloguing tool was developed in the Mkbeem project to support publishing of monolingual product information in multiple languages in a write-once-publish-many manner. The cataloguing constitutes the central back-office routine of the Mkbeem system and prepares information that is used by the real-time customer service, which is the other major part of the Mkbeem system.

Cataloguing of a new product goes through multiple steps, taking into account linguistic, product model and culture-specific issues. Figure 2 illustrates the user interface of the cataloguing tool. The cataloguing starts by importing or typing the new product article into the left-most pane and importing the corresponding product photograph to its place. After finishing with the text, it is important that the user checks it by selecting <Check_Text>. The system verifies that the description article conforms to the language model in order to guarantee the quality of the following automatic processing steps and proposes corrections when necessary, e.g. removing ambiguities

When the product description text has been successfully verified, the user may continue the cataloguing process. Selecting <Extract_Properties> activates the next step**.** This functionality finds product properties from the textual product descriptions along the provided ontological product model (detailed description in [10]). The extracted properties are language neutral and they are stored in a relational database for further uses in inference, product classification and natural language information request processing. Inference means that new information is concluded based on related generic ontologies. E.g., for a given cloth qualitative facts are inferred based on a material ontology and the known material composition of the cloth. The system can also find out origin (e.g. are they natural or synthetic) of cloth materials. Based on a generic colour ontology colour similarities and harmonies can be inferred [11]. The results of the property extraction are shown to the user in the right-most pane as a property tree.

Selecting <Find_Categories> derives a proposal for categorisation of the product with respect to predefined market-specific product hierarchies. For each target locale there may be defined product hierarchies of its own. The goal is to find all possible categories with which the product can be justifiably associated. In many cases several categories are associated with a product thus providing several redundant paths for the customer to locate it. The categorisation is based on the extracted properties and simple rules. Localisation-specific issues arise, e.g., in the case of winter clothes as the conception of "a winter cloth"

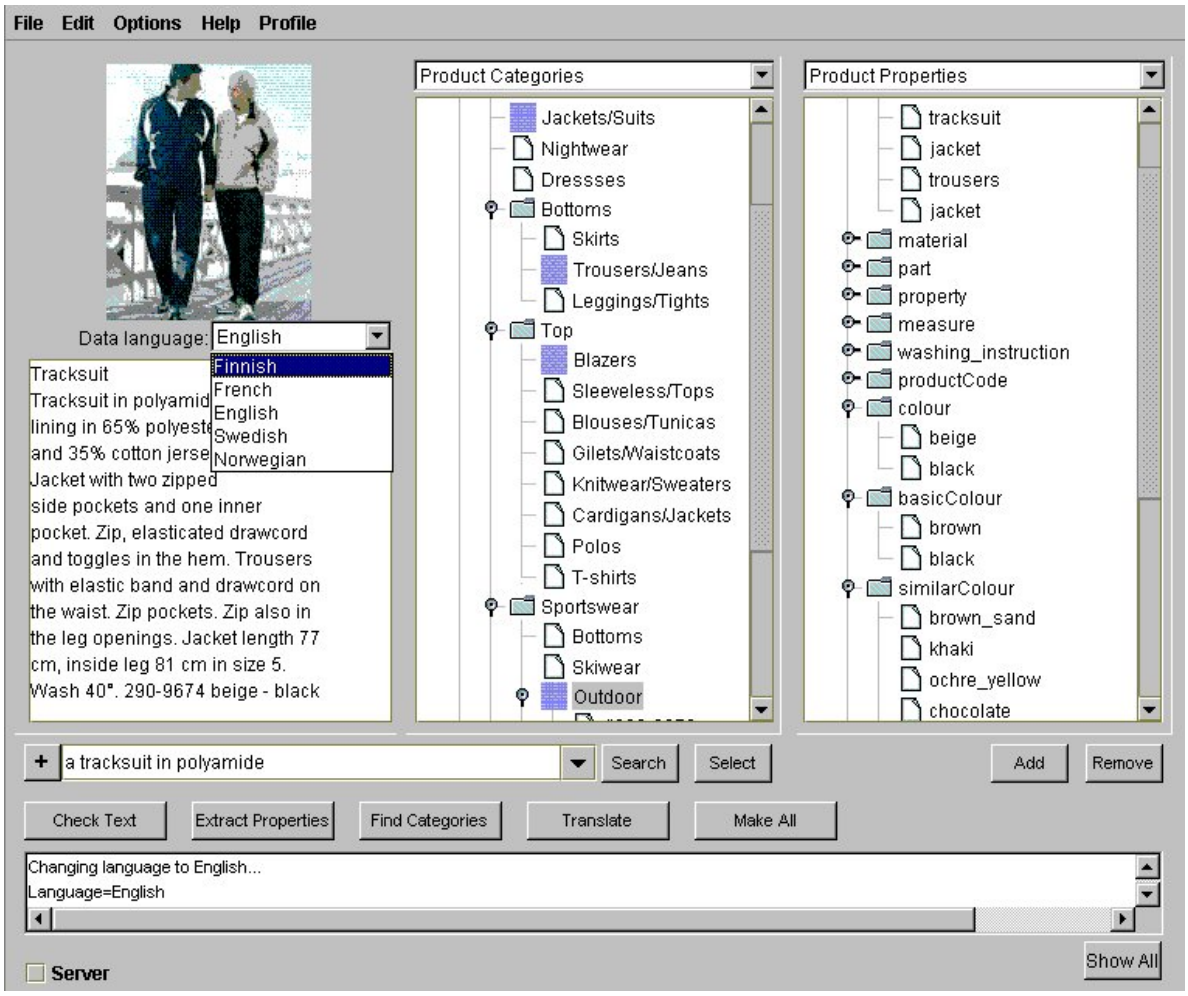

Figure 2. User interface of the multilingual cataloguing tool.

differs in Finland and in Greece. The categorisation results are presented in the middle pane. The user may override the suggestions of the system.

The next step involves fully automatic translation of the product description into all covered target languages by selecting <Translate>. The preceding text checking has ensured that the source text does not contain any parts that cannot be automatically translated. The verified product article is passed to the Webtran sublanguage machine translation system [4,7,8,9] to get accurate translations that need none or just minimal post editing. Webtran uses translation grammars that are quite closely related to the product ontologies. The translation process is partly interlingual and partly transfer based, as some language expressions do not have equivalents in the ontologies. Translation grammars are best defined at the same time with the product ontologies. If sample translations are provided, a machine learning algorithm can be used for deriving a translation grammar [15]. In principle, the Webtran translation technology is language independent. In Mkbeem test pilot it was used to translate in the following directions: Finnish->English, Finnish->French, French->English, Swedish->Finnish, Swedish->English. When the translation is done in the cataloguing tool, the user may change the data language view from the combo box below the photograph and go to check the translation results and find the market specific categories for the product.

In order to support the cross-checking, browsing and maintenance of the product selection, the cataloguing tool includes a natural language information request facility. It lets the user to type information requests using noun phrases like “a brown tracksuit in polyamide”. The meaning of the query is analysed with respect to the product ontologies

and then the recognised constraining properties are matched against saved properties of the products in the database.

After the previous steps the product information is stored to the product database. This information includes the translated product articles, the extracted properties, the deduced additional properties, and the market specific categories where the product belongs.

5. User Experiences

The testing of the cataloguing tool was carried out by the mail-order company Ellos Postimyynti Oy. The idea of the testing was to test the concept of a cataloguing tool, which supports the maintenance of a multilingual e-commerce product catalogue using a pivot language and a single tool. The usability of the cataloguing tool was tested in a real working environment in the autumn 2002. The test group consisted of eight persons, who were translators, cataloguers or call-centre workers. At the time of the tests the language model for Finnish, French and English consisted of about 815 syntax and transfer rules, 150 meaning extraction rules and the lexicons consisted of about 11000 lexical items. Product models included 62 concepts.

The testers were interviewed before the trials to record their background, experience with computers and expectations. Six testers told us being familiar with catalogue maintenance. All mastered Finnish, six English, four Swedish, and one French. All were in some degree familiar with computers and the Internet. They were also given guidance of the use of the cataloguing tool. After a test period of about one month, there was carried out a second round of interviews, were the experiences were recorded.

To the question “How do you judge each main feature of the prototype you tested (good or insufficient)?” we got the answer of Figure 3. Every cataloguing professional had answered, but not all of the call-centre workers.

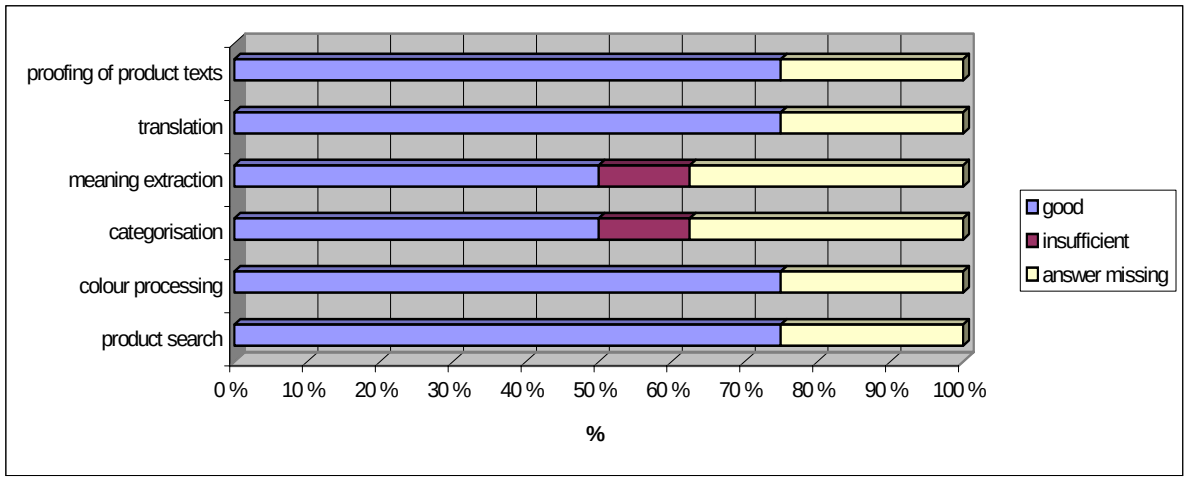

Figure 3. Test user evaluation of the cataloguing tool functionalities.

Some of the answers were very interesting. For instance for the question “Were you surprised by the category Mkbeem proposed you for cataloguing an offer?” came the following answers Yes (4), No (1), answer missing (3), while for the question “Were these proposals wise (whether surprising or not)?” Yes (4), answer missing (4). From this we can conclude that the automatic proposing of suitable categories works generally fine and brings added value by finding valid categories that cataloguing professionals would have otherwise omitted totally.

Test-users consider the cataloguing tool to be a useful tool for the production of multilingual product information, and each of the main features of the prototype was reported as good. Besides the very important possibility of semi-automatic translation into

target languages, test-users named functionalities like property extraction and inference with colours and materials to be very interesting. It was found to be important in bringing the customer new possibilities to find accurate information from diverse sources.

The cataloguing process as a whole was seen as an easy and efficient way of producing and classifying product information. One important advantage in an integrated cataloguing environment is that it helps in producing consistent and uniform information as the whole cataloguing process is based on joint language and product models that conform to the company knowledge of the domain. Moreover, test-users anticipated that the use of the cataloguing tool can make the working process faster and it reduces the amount of manual, repeated routine procedures.

The machine translation component Webtran of the cataloguing tool has been in production use at Ellos since the year 2000. The EUROMAP case study done by CSC Inc. [13] reports savings of over 30% in translation time having been reached after a relatively short use of this component. Currently, it translates around 15.000 product articles annually. Its users consider the component nowadays crucial for their work.

6. Business Benefits

The cataloguing tool accelerates the preparation of multilingual product information and reduces the time-to-market for new products. By lowering the information expense per product and per language, the tool lets cost effectively expand the product selection of eShops. The end-users gain access to a wider range of products in their native languages.

In order to test how easy is to adapt the cataloguing tool to other domains of goods than clothing products, we adapted the tool for cataloguing vacation cottage information to mediate Finnish cottages to French and English clientele. Based on our modelling experiences we have estimated that introducing a comparable new domain would require the following work amounts: building the semantic-lexicon 2 man-months, making translation and meaning extraction rules 1 man-month and creating the product models 2-4 man-weeks.

We also estimate that adding a totally new language to a pre-existing domain would need the following work: acquisition of semantic-lexicon 1-2 man-months, defining translation and meaning extraction rules 2-4 man-weeks, adjustment of product models 1 man-week.

At least for domains of relatively large customer interest and trading volume, the marginal cost of adding a new domain or a new language is reasonable with respect to the added value gained. In the future, we plan to develop tools that would accelerate the adaptation using machine learning techniques.

7. Conclusions

This paper described the multilingual cataloguing tool that was developed for the Mkbeem mediation system. The methodology, namely combined use of ontologies and language processing, was outlined. The basic functionalities of the tool were described. Finally, user experiences from the field tests of the cataloguing tool were summarised.

Currently the tool suits well for the production of paper or web-catalogues for international markets. On the basis of the test results from the end-user trial, further development of the cataloguing tool in future projects could focus on the following issues:

- Further developing the information request processing. For instance null queries should be possible to avoid using an ontology based attribute constraint relaxation strategy. Also the dialogue skills of the system would be good to further develop.
- Adding new languages to the system. For instance German, Norwegian, Russian and Estonian have been mentioned in the discussions with the testers.

- More user-friendly ways for the acquisition and maintenance of language models and product models are needed. For instance the knowledge acquisition processes could be partially automated by using machine learning techniques. The overall goal is to make the adaptation of the Mkbeem technology to new product domains and new languages affordable and easy to even small businesses.
- In the future, the special needs of other media than paper or web should be taken into account, such as mCommerce with PDAs or mobile phones. This would also require further development of natural language processing capabilities, like automatic text abstraction for textual product descriptions to support smaller screen size.

There are two ways of getting the cataloguing tool technology into production use. Firstly, a portal operator could provide the tool for particular domains. This could be an easy solution for domains with large number of small or even micro businesses involved, like renting of vacation cottages. Among the project partners one is planning to utilise the technology for their catalogue services and another for travel services. Secondly, the tool can be embedded into the catalogue production process of a seller company. It likely changes the organisation of work and requires business process re-engineering. The cataloguing tool can not be considered as package software. Implementing it to production use requires adaptation and tailoring. Currently, there are negotiations going on with a large European trading house about exploitation of certain components of the cataloguing tool.

Test experiences have clearly shown the need for this kind of multilingual cataloguing tool in the every day business of a company operating in multilingual, multicultural and multinational markets. We are looking forward to implementing it into commercial use in companies and to developing new innovative features into it in our future research projects.

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

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

[2] GlobalReach (2002): Statistics on the website of Global Reach Inc, May 2003, URL http://www.glreach.com/

[3] Gomez-Perez, A.; Corcho Carcia, O.; Fernandez Lopez, M.; Lehtola, A.; Taveter, K.; Sorva, J.; Käpylä, T.; Toumani, F.; Soualmia, L.; Barboux, C.; Castro, E.; Sallatin, J.; Arbant, G.; Bonnaric, A. (2001): Requirement, Choice of a Knowledge Representation and Tools. Public Report of MKBEEM project (EC IST-1999-10589), Version 2.0, available e.g. from www.mkbeem.com, 93 p.

[4] Jaaranen, K.; Lehtola, A.; Tenni, J.; Bounsaythip, C. (2000): Webtran tools for in-company language support. In: Language Technologies for Dynamic Business in the Age of the Media. Köln, 23 - 25 Nov. 2000. Vereinigung für Sprache und Wirtschaft. Köln (2000), pp. 145 - 155.

[5] Leger, A.; Lehtola, A.; Villagra, V. (2000): MKBEEM – Developing Multilingual Knowledge-Based Marketplace. ERCIM News, July 2001, pp. 50-52. Reprinted in Research News of VTT Information Technology, December 2001, pp. 1 – 3.

[6] Leger, A.; Michel, G.; Barrett, P.; Gitton, S.; Gomez-Pere, A.; Lehtola, A.; Mokkila, K.; Rodrigez, S.; Sallantin, J.; Varvarigou, T.; Vinesse, J. (2003): Ontology domain modeling support for multi-lingual services in E-Commerce: MKBEEM. ECAI'00, Workshop on Applications of Ontologies and Problem-Solving Methods. Berlin, 20 - 25 August 2000. Berlin, 4 p. URL http://delicias.dia.fi.upm.es/WORKSHOP/ECAI00/19.pdf

[7] Lehtola A.; Tenni J.; Bounsaythip C.; and Jaaranen K. (1999): WEBTRAN: A Controlled Language Machine Translation System for Building Multilingual Services on Internet. In proceedings of: Machine Translation Summit VII `99 (MT Summit 99), September 13-17, 1999, Singapore, pp. 487 - 495.
[8] Lehtola, A.; Tenni, J.; Bounsaythip, C.; and Jaaranen, K. (1999): Controlled Languages as the Basis for Multilingual Catalogues on the WWW. In: Business and Work in the Information Society: New Technologies and Applications, Roger, Stanford-Smith and Kidd (Eds.). IOS-Press, Amsterdam, pp. 207-213.
[9] Lehtola, A.; Tenni, J.; Taveter, K.; Käpylä, T.; Silvonen, P.; Jaaranen, K. (2001): Engineering Multilingual Internet Commerce. In: C. Stephanidis (ed), Universal Access in HCI, Vol. 3 of the Proceedings of HCI International 2001, Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, pp. 501 - 505.
[10] Lehtola, A.; Tenni, J.; Käpylä, T. (2003): Multilingual Cataloguing of Product Information of Specific Domains: Case Mkbeem System. In: Proceedings of EAMT/CLAW´03, Dublin, May 15-17, pp 79-85.
[11] Lehtola, A.; Heinecke, J.; Bounsaythip, C (2003): Intelligent Human Language Query Processing in Mkbeem. In: C. Stephanidis (ed), Universal Access in HCI, Vol. 4 of the Proceedings of HCI International 2003, Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, pp. 750-754.
[12] Levy, A.; Rousset, M.C. (1998): CARIN: A Representation Language Combining Horn rules and Description Logics. Artificial Intelligence Journal, vol. 104, September 1998.
[13] Loimaranta, O. (2000): EUROMAP HLT Case Study: Webtran – a controlled language machine translation system for building multilingual services on Internet. December 2000, http://www.hltcentral.org/usr_docs/case_studies/euromap/FIN_webtrans.doc
[14] Mkbeem (2000-2002): The MKBEEM project site. URL http://www.mkbeem.com/
[15] Tenni, J.; Lehtola, A.; Bounsaythip, C.; Jaaranen, K. (1999): Machine Learning of Language Translation Rules. In proceedings of: 1999 IEEE Systems, Man and Cybernetics Conference (SMC'99), October 12-15, Tokyo, pp. 171-177.