In: C. Stephanidis (ed), Universal Access in HCI, Vol. 3 of the Proceedings of HCI International 2001, Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, 2003, pp. 501 - 505.

ENGINEERING MULTILINGUAL INTERNET COMMERCE

Lehtola, A., Tenni, J., Taveter, K., Käpylä, T., Silvonen, P., Jaaranen, K.

VTT Information Technology
P.O.Box 1201, FIN-02044 VTT, FINLAND
email: aarno.lehtola@vtt.fi

In the paper we describe our three tools that address the WWW multilinguality, namely Webtran, CONE and Unifier software. Webtran is a machine translation software that has been specifically designed for fully automatic translation of domain specific texts in online services. It is in daily use translating product descriptions in a mail-order company in Finland. CONE is a conceptual network software that includes an editing tool for defining domain ontologies embedding linguistic information and a user interface tool enabling multilingual navigation to information. It has been used to provide cross-lingual access to consumer protection legislation of some countries in European Union (EU). Unifier software helps in correcting erroneous or inexact text inputs of end-user customers in online services. For instance, it can be embedded to correct and homogenise address information.

1. Introduction

B2B electronic commerce has been estimated by Gartner Group to grow at over 40 % a year for the following five years (Tieke 2000). Even more enthusiastic growth figures have been presented. For instance, Boston Consulting Group has forecast that Internet commerce now triples in each forthcoming year. This growth takes place in tandem with the liberation of the world trade and the formation or expansion of common markets. While the trade transactions more and more frequently cross country and language borders, there is an extensive need for technical solutions to make the e-commerce services multilingual. VTT responds to these challenges by developing tools for building multilingual, localized services on the internet. In the following we review some central requirements that a multilingual e-commerce service should fulfil.

High quality of user interface localisation will be crucial in building positive image of an e-commerce service. Especially in the consumer markets, there will be many users that are illiterate in any other language than their native language. That is why multilinguality must be implemented in a consistent and comprehensive manner. McKinsey & Company has found in a survey that from all visitors of e-commerce sites (over 1.8 million visitors considered in the survey) only 7 % became customers and only 1,3 % became repeat customers. According to a survey by Jupiter Communications 39 % of consumers are more likely to shop online at merchants with whom they have had an offline experience (Ramsey 2000). There are plenty of inexperienced computer users among the clientele, so the user interface must be as simple as possible to use by a customer. It is crucial for an e-commerce service to get a high-quality image starting from the first access by a new customer.

Adaptation to the circumstances of the target market area is important. A seller may have some external reasons, why the product repertoire is not the same in every country. E.g. the seller may have rights to sell in only some countries, some items may have luxury taxes etc. At latest, when a customer has selected some product, there must be shown: price in local currency, methods of payment, methods and costs of delivery, delivery time and other terms of delivery. As far as the transactions cross over borders of the EU also taxation/customs needs to be considered. There are issues that depend on local legislation, such as returning and refunding of the goods. The multilingually provided information must be reliable to avoid legal disputes. The used language adaptation techniques must guarantee accuracy.

E-commerce services should be quick enough to prevent the customer from changing the channel. In a survey by Service Metrics Inc. (U.S.) it was found that the average download time for e-commerce sites is between six and seven seconds. It was also recognized that most customers will get annoyed if forced to wait more than eight seconds for a downloading of pages. There are several surveys on the percentage of aborted trading sessions, i.e. "abandoned shopping carts". For instance, both Andersen Consulting and Forrester Research have resulted with 25 % of all sessions, eMarketer with 31 %, Visa with 43 % and Greenfield Online with even 67 % (Ramsey 2000). Quick response times are of essence for customer satisfaction.

Additional requirements include support of a wide range of terminals, high security requirements, good integration to legacy systems, portability and homogeneity of the solution, and scalability and expandability.

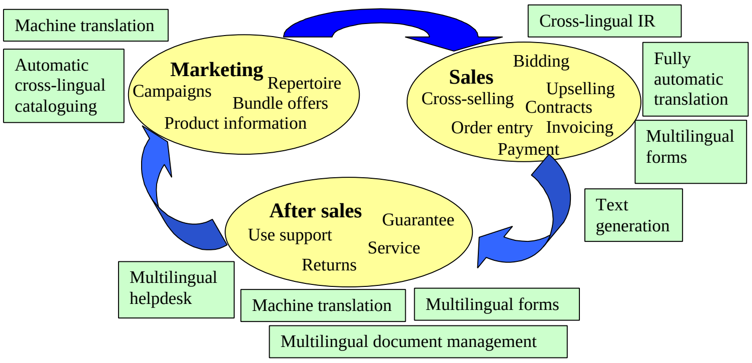


Figure 1. E-commerce processes in company-to-customer interface.

Figure 1 outlines the processes, that are prevalent in company-to-customer interfaces. In the boxes are presented the technical solutions that provide multilinguality. Cross-lingual information retrieval (IR) may involve correcting and unifying input search terms, ontology based translation of the query and finally translating the search results. Multilingual forms adapt to locale by their graphical appearance (layout, colours, icons, text orientation and direction, sounds etc.), translated fixed texts and processed form data. The last one involves, e.g., correction and unification of user inputs, text generation and machine translation. Multilingual helpdesk could be based on a database of centralised questions and answers to be maintained in one pivot language, and cross-lingual IR. Automatic cross-lingual cataloguing finds for a new product the suitable categories from locale specific catalogues. These examples involve functions that can be implemented with our software tools Webtran, CONE and Unifier.

2. Automatic translation of in-company language

Webtran software (Jaaranen 2000, Lehtola 1999a, Lehtola 1999b) is a tool for authoring and automatic translation of domain specific texts that are written in a human sublanguage characterised by selected domain, vocabulary and sentence structure, such as technical documentation: manuals and product descriptions, and formal reports: e.g. weather forecasts, medicine effect descriptions and epicrises. Webtran is currently in production use in the mail-order company Ellos Postimyynti Oy in Finland for automatic translation of the Ellos sales catalogues from Swedish into Finnish. Automatic translation shortens the time-to-market of products and improves the overall competitiveness of the company.

Figure 2 below illustrates an arrangement with Webtran translating fully automatically texts of a repository. The figure subsumes that the original texts have been authored and edited in a pivot language with controlled conformance to the in-company language model and its constraints, and the texts therefore include a minimised amount of ambiguities. The in-company language model is extended and maintened using the Webtran language modelling tool. The translation engine can automatically and accurately translate approved texts to multiple target languages. Major savings are obtained as post-editing of the translations can be avoided.

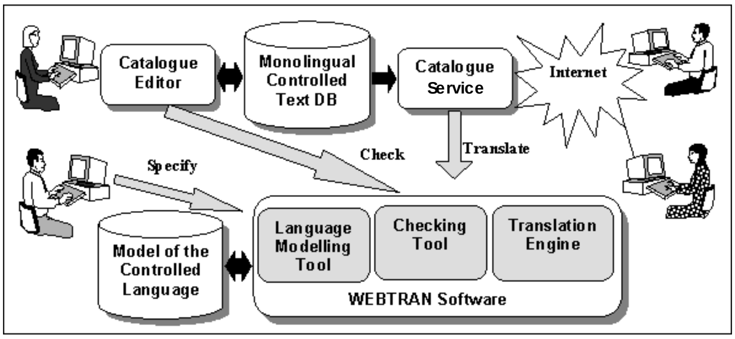


Figure 2. Embedding fully automatic translation into an online service.

3. Cross-lingual information retrieval using ontologies

CONE software for conceptual network modelling and reasoning (Taveter 1999) is a tool for supporting cross-lingual information retrieval, building of market specific product models (ontologies), and data mining for e-commerce. CONE enables the modelling of concepts and interconceptual relationships at different logical levels. The models also include linguistic knowledge: the ALE-rules (Augmented Lexical Entries) (Lehtola 1999b) attached to the concepts and relations conduct the recognition of different ontological constructions that correspond to certain expressions in natural language, and the generation of textual expressions on the basis of ontologies. In other words, the use of linguistic information combined with ontologies enables the interaction between the user and the online service provider in human language, thus making virtual shopping experience more natural. Use of ontological approach also enables a more efficient way for the customer to make searches and to find relevant products in online product catalogues than with a normal keyword–based search.

As Figure 3 reflects, linguistic information in form of terms in different languages is expressed as concept properties denoted by the language abbreviations in parenthesis. Each term is further characterized by a fuzzy value, expressing the degree of exactness with which the term confirms to the concept. For instance, the utterance "*Two-piece jogging outfit with elastics at the leg cuffs*" might be a part of the user dialog in an e-shop 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 recognized. Relations between the concepts are also recognized by using the ALE's shown beside the relations in Figure 3.

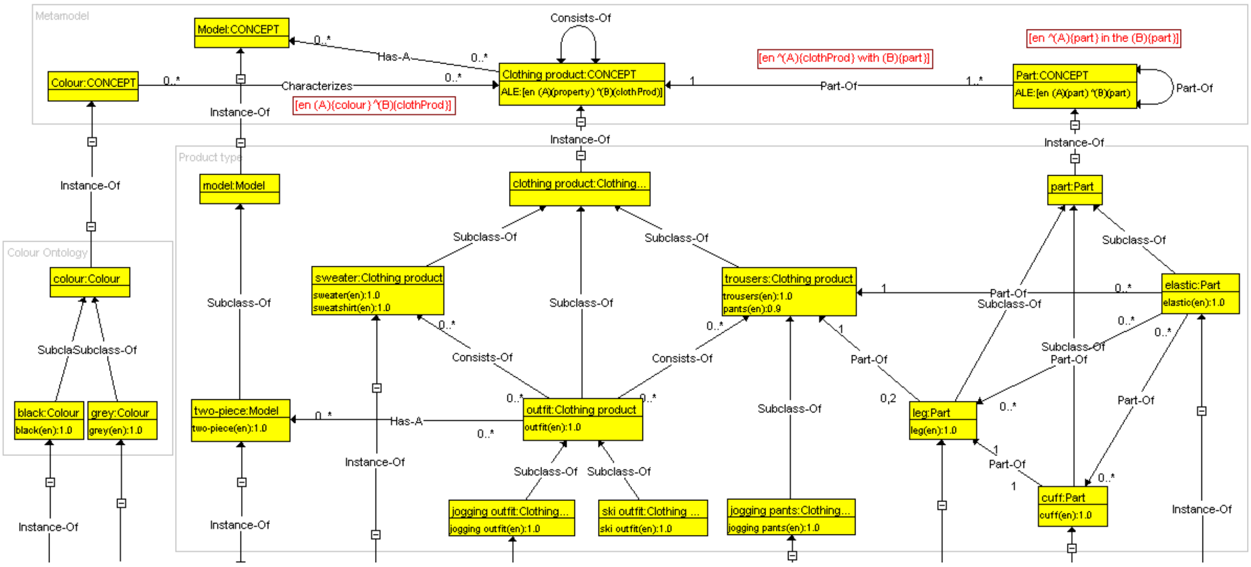


Figure 3. An excerpt from an ontology covering clothing products

In addition to identifying product articles based on indistinct user utterances, ontological approach of the described kind also enables online upselling (offering a bit more expensive and better product article) and cross-selling (offering additional product articles based on the user's selection). Our approach also helps in finding suitable products to be bundled for offering and selling together. CONE has also been used for ontology-based translation of query terms in a legislative domain. Here, the conceptual approach is preferred over traditional thesaurus because translation and use of legal terms depends on the differences between law systems which can be described by ontologies.

4. Correcting and unifying input texts

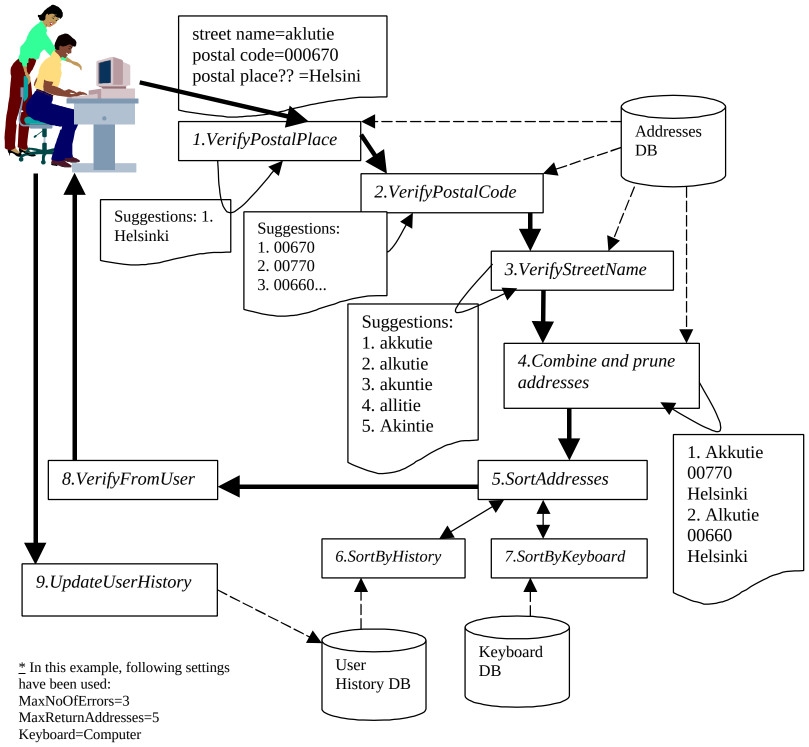
Unifier software helps in correcting and unifying erroneous or inexact text inputs of online customers in e.g. search engines or order and registration forms. It also facilitates the management of customer information, e.g., in order processing and in data mining. The software also applies to correction of inexact database material collected, for example, by optical character recognition (OCR) devices. Databases often contain many variations of the product names, only slightly modified. It is easier to utilise this information, if the used terminology is analogous; hence the products are easier to be categorised according to their names, e.g. for auditing purposes.

Unifier checks and corrects errors both on word and phrase level. Phrase level error detection is integrated with Webtran language modelling tool described above. It is used to guide the user in word and sentence selections to ensure that the generated text confirms to the definition of a domain specific language model. Word-level errors are detected with dictionary lookup. The dictionary structure can also be used as a thesaurus, which is helpful in applications where it is important to homogenise the terminology. For example, different variations of product names and features can be declared as synonyms. This improves the use of search engines, as different variations produce the same search result.

Unifier can make use of domain knowledge, which enables better accuracy and performance. The contextual information can be used for dictionary partitioning: in order forms the possible values for a field can be declared in a special dictionary. Different dictionaries can be attached to each other: e.g., when checking addresses, the postal codes can be attached to post office names to ensure their mutual accordance.

Unifier ranks the correction candidates according to keyboard neighbourhood. Different keyboard structures can easily be added to the system. Unifier adapts to user's misspelling habits, which is done by collecting user history statistics. The history is used in ranking of the correction methods.

Unifier has been piloted in address correction, which can be embedded in e-commerce order forms. The following Figure 4 shows the process of postal address verification and correction in online services. The process starts when a user enters postal address in online form (e.g. order form, registration form) and clicks on "Submit"-button. Unifier first verifies the postal place (phase 1) and finds out that there is a typo in it and thus returns the only correction suggestion resembling the user entry. At the next phase (phase 2), Unifier verifies the postal code. As several postal codes match the postal place they all are returned. At the phase 3, the street name is verified. Again, there are multiple correction suggestions, of which 5 first are returned (as 5 is here the maximum number of return values). At the phase 4, street names are combined with 'postal code-postal place' -tuples. At this time, impossible combinations are discarded. The two satisfactory entries are sorted, first using information of typical error types found from user history and then according to the keyboard. The correct address entries are shown to the user in probability order. The user selects the preferred one, and the selection is then used to update the user history.



* In this example, following settings have been used:
MaxNoOfErrors=3
MaxReturnAddresses=5
Keyboard=Computer

Figure 4: Practical arrangement for correcting postal addresses in online services

5. Conclusions

Webtran, CONE and Unifier are language independent. Tests have been carried out with English, French, Swedish, Estonian and Finnish. In the future, the number of languages will be increased. Currently, VTT is participating in the EU-project Mkbeem (Multilingual Knowledge Based European Electronic Market place, 2000-2002) where a mediation system is developed to enable the access to online products and services in the native languages of customers (Léger 2000). Mkbeem explores new ways of combining human language processing with ontologies for realising three very basic use scenarios of a multilingual online shop: multilingual cataloguing, multilingual information retrieval, and multilingual trading. In MKBEEM, the semantics of the products, i.e., their properties and internal relationships, are desribed in domain ontologies whereas the the content provider ontologies describe product categories of a product catalogue (Gómez-Pérez 2001). These semantics are used to cope with human language ambiguities in order to provide efficient and accurate human language processing. As a proof of concept, the project is implementing a mediation service which supports three European languages: French, English and Finnish.

References

Gómez-Pérez A., Corcho García O., Fernández López M., Lehtola A., Taveter K. , Sorva J., Käpylä T., Tourmani F., Soualmia L., Barboux C. , Castro E., Sallantin J., Arbant G., Bonnaric A. (2001). Requirements, Choice of a Knowledge Representation and Tools. 126 p. http://mkbeem.elibel.tm.fr/

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

Léger, A., Michel, G., Barrett, P., Gitton, S., Gomez-Pere, A., Lehtola, A.., Mokkila, K., Rodrigez, S., Sallantin, J., Varvarigou, T., & Vinesse, J. (2000). Ontology domain modeling suppor 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., Tenni, J., Bounsaythip, C., & Jaaranen K. (1999a). Controlled Languages as the Basis for Multilingual Catalogues on the WWW. Jean-Yves Roger, Brian Stanford-Smith and Paul T. Kidd (Eds.). *Business and Work in the Information Society: New Technologies and Applications.* IOS-Press, Amsterdam, pp. 207-213.

Lehtola, A., Tenni, J., Bounsaythip, C., & Jaaranen, K. (1999b). WEBTRAN: A Controlled Language Machine Translation System for Building Multilingual Services on Internet. In: *Proceedings of Machine Translation Summit VII `99 (MT Summit 99)*, Singapore, pp. 487 - 495.

Ramsey, G. (2000 March). Online sales surged over the holidays. *Business 2.0 Magazine*, pp. 433-434.

Tieke (2000). Edisty Journal, No. 4, *Finnish Information Technology Development Centre*.

Taveter, K., Lehtola, A., Jaaranen, K., Sorva, J., Bounsaythip, C (1999). Ontology-Based Query Translation For Legislative Information Retrieval. In: *Proceedings of the 5th ERCIM Workshop On User Interfaces For All (UI4All)*, Dagstuhl, Germany, 1999, pp. 47-58.