Report on the MULSAIC'96 Workshop
"Multilinguality in Software Industry: the AI Contribution"

Constantine D. Spyropoulos and Vangelis Karkaletsis

National Centre for Scientific Research (N.C.S.R.) "Demokritos",
Institute of Informatics & Telecommunications, Aghia Paraskevi 15310, Athens, Greece
Tel: +301-6513110 ext. 520, Fax: +301-6532175
E-mail: {costass, vangelis}@iit.nrcps.ariadne-t.gr

Abstract

This is a report on the MULSAIC'96 Workshop that was dedicated to examine the contribution of AI methods, techniques and tools to support multilinguality in software industry. Here we present the main topics covered during the workshop and give short descriptions of the contributed papers.

Introduction

Software development, typically, is oriented towards a primary market, i.e. the software "speaks" the language of this market, follows its rules, conventions and customs, and does whatever is necessary to make users from this market feel comfortable. Any installation in other markets, with different languages and culture requires either that the users adapt themselves to the product, using its language and conventions about cultural aspects, or that the product be converted for use in the local market.

With the growing PC user community more and more people request, and rightly expect that the software they use talks their language and respects their cultural background. Furthermore, with the software market expanding into new uncharted territories, like the countries of the former Soviet Union, China and the countries of Africa it is to be expected that a large percentage of the new users will not be able to speak other languages or understand other cultural conventions than their own. Even in Western European countries, like France, legal restrictions are imposed onto imported software, at least as far as language is concerned. Additionally, more and more international companies want to use existing software in their various subsidiaries worldwide. The investment in the software typically is so expensive that repeated development for the various countries is not justified anymore. They require globalized software.

Therefore, multilinguality in the software industry is a major demand nowadays for the companies that want to make business internationally or have subsidiaries worldwide.

The question that appears is whether AI based methods and tools can be used by software industry to support the development of multilingual software. Although a variety of workshops or conference sessions have already tackled problems concerning multilinguality (i.e. TKE, IJCAI PRICAI etc.), MULSAIC'96 was the first attempt focusing on how the AI can contribute to help software industry in order to solve problems concerning multilinguality.

The workshop was held during the 12th European Conference on AI (ECAI-96), Budapest, 11-16 August 1996, and brought together people from the software industry and the AI research community. It provided an opportunity for people from the software industry to present particular problems and questions associated with multilinguality as well as an opportunity for AI researchers to show that AI-based methods and tools can help improve the current situation. The successful organisation of MULSAIC'96 and the strong interest of the participants for another opportunity to discuss multilinguality and the AI contribution motivated us to organise the MULSAIC'97 Workshop in the context of 15th International Joint Conference on AI (IJCAI-97), Nagoya, Japan, August 23 - 29, 1997.

MULSAIC'96 was organised in four sessions. The topics addressed in each session and short descriptions of the papers presented there are discussed in the next section. Some information on MULSAIC'97 is given in the final section.

Contributions to MULSAIC'96

The papers submitted in MULSAIC'96 were 13 in total. From these papers, 10 were presented during the workshop. Postscript versions of all the papers can be found in the WWW page:
http://www.iit.nrcps.ariadne-t.gr/~costass/mulsaic.html.

Session 1: Software Companies Experiences in Multilinguality

This session addressed the problems software companies face during the development of multilingual software

applications as well as the strategies some of them adopt to handle these problems. Problems that multilinguality imposes may be

- exchanging data in different languages between countries,
- supporting different languages in the user interface,
- processing multilingual documents,
- retrieving multilingual information,
- ensuring consistency in terminology and in writing and translation style,
- reducing the translation and in general the localisation cost.

The two papers presented in Session 1 mainly discuss the experiences of two large multinational companies, Lotus and Sofware AG, concerning multilinguality:

Alan Barrett, in "The Lotus Multilingual Strategy", concentrates on the main components of the Lotus multilingual strategy and presents the "Notes Global Designer", a set of tools that allow the developer to generate multilingual applications using standard terminology.

Andreas Schütz, in "Software AG Experiences in Producing Multilingual Software", concentrates on the multilinguality problems Software AG developers and localisers face during the localisation of their products (inconsistency in terminology, uncontrolled technical writing, differences in space required by different languages, multilingual text processing difficulties), discusses the main components of Software AG strategy in handling these problems and presents the results of investigations on the use of language technology tools, performed by Sofware AG.

Session 2: Multilingual Resources and Intelligent Translation Tools

The papers of this category address the contribution of AI methods and techniques in the development, management and exploitation of multilingual resources (lexicons, thesauri, terminological bases, grammars, ontologies, translation bases) and translation tools (translation memory, translation editors, terminology management systems, translation workbenches, machine translation). Such contribution mainly involves the

- automatic acquisition of multilingual resources,
- efficient organisation and management of multilingual resources,
- exploitation of multilingual resources in software localisation and internationalisation,
- development of intelligent translation tools.

The five papers of Session 2 concentrated on:

- the automatic extraction of translation equivalents from parallel texts,

- the organisation of terminological and translation resources,
- the mapping of word meanings from different languages onto a core concept ontology,
- the development of intelligent translation tools.

More specifically:

Sotiris Boutsis and Stelios Piperidis in "Automatic extraction of bi-lingual lexical equivalencies from parallel corpora" propose the use of a method that relies on statistical and corpus-based techniques for the automatic extraction of translation equivalents. The use of such techniques ensures the easy application of the method in different language pairs.

Alexander Geyken and Vladimir Tourovski in "Extraction of example phrases for a bilingual dictionary of verbs" propose the use of a frame-based bilingual (French-German) lexicon that contains semantic information (simple selectional restrictions of the form "animate-inanimate") and a simple algorithm that makes use of it

Vangelis Karkaletsis and Costas Spyropoulos in "Exploiting an AI-based approach in multilingual software industry" propose the use of a terminological knowledge base and a knowledge representation system for organisation of terminological knowledge in multilingual software products.

Bróna Collins and Pádraig Cunningham in "Translating software documentation by example" present a generic frame-management system, for the storage of a translation base. Each entry in this base can be indexed by entire strings, individual words or syntactic functions.

Richard Sutcliffe, Annette McElligott, Donie O Sullivan, Anatoliy Anatolievich Polikarpov, Leonid Aleksejevich Kuzmin, Gearóid O Néill and Jean Véronis in "An Interactive Approach to the Creation of a Multilingual Concept Ontology for Language Engineering" investigate methods by which word senses in a variety of different languages can be mapped onto the core concept ontology of the International WordNet project.

Session 3: Multilingual Information Services

The contribution of AI methods, techniques and tools in the area of multilingual information services was also addressed. This contribution mainly concerns

- multilingual information retrieval,
- intelligent information services,
- multilingual information exchange.

The four papers of this session concentrated on:

- the retrieval of information written in a language different than the one used in the user's query,

- the presentation of the information retrieved into the user's (native or selected) language,
- the selective access of information written in a language specified by the user,
- the interpretation of the available information according to the user requirements and the presentation of the results to the user in his native or selected language,
- the use of language technology to model electronic bargaining in order to support multilinguality

More specifically:

Genichiro Kikui, Yoshihiko Hayashi and Seiji Suzaki. in "Cross-lingual Information Retrieval on the WWW" present a cross-lingual search system they have developed that provides interface in the native language while performing multilingual searches on the WWW. This system relies on two AI modules: a language identification module based on statistic and rule-based pattern matching, and a quick MT module.

Christian Fluhr, Dominique Schmit, Philippe Ortet, Karine Gurtner, Vera Semanova and F. Elkateb in "Distributed multilingual information retrieval" present the approach of the EMIR project in multilingual information retrieval. They discuss the use of a "reformulation" module that can infer new words from the original query words according to a lexical semantic knowledge base (reformulation rules). They also discuss the problem of managing databases containing documents in several different languages and propose an approach for solving it.

Jörg Schütz in "Intelligent Web-based Information Services" presents the AI techniques employed in an on-line multilingual technical information service system of the automotive industry. He proposes the use of a knowledge-based representation of domain, product and production knowledge. This knowledge base contains "language-independent" concepts under which domain specific terminology dictionaries are classified.

Aarno Lehtola and Timo Honkela in "Multilinguality in Electronic Commerce - Research Issues" raise the problems of multilinguality that have to be solved to support the International Electronic Commerce and they propose ways on how NLP, Machine Translation and Artificial Neural Networks can help to overcome these problems.

Session 4: Multilingual Generation of Messages and Documentation

The contribution of AI methods, techniques and tools in the area of multilingual generation of messages and documentation in software products was addressed in this session. This contribution mainly concerns

- the on-line and dynamic generation of multilingual messages
- the automatic generation of draft documentation
- the support of technical writers and translators in writing messages and documentation,

and was expressed by the two papers of the session:

Serge Sharoff and Lena Sokolova, in "Contrastive analysis of software manuals", discuss cultural phenomena and linguistic considerations for manuals in Russian and English while looking to several phenomena in German. The features that are valid across languages are investigated and compared to other types of technical manuals. Finally, a contribution of the proposed analysis to the multilingual generation of software manuals is discussed.

Costas Spyropoulos and Vangelis Karkaletsis, in "Online Generation of Messages: A Knowledge-based Approach", propose a knowledge-based approach for the online and dynamic generation of multilingual and user-tailored messages. The benefits and costs of this approach are discussed and the results of a prototype development for the generation of bilingual (English, Greek) diagnostic messages in a software product are presented.

MULSAIC'97

MULSAIC'97, based on the results of MULSAIC'96 intends to promote the discussion between people from software industry and AI in issues such as the following:

- Multiscript communication through networks
- Localisation of user interfaces
- Cross-language, language-independent text retrieval
- Language identification techniques
- Multilingual text generation
- Multilingual information extraction
- Machine Translation (MT) and Machine-Assisted Human Translation (MAHT)
- Creation and use of multilingual resources (dictionaries, terminologies, corpora)

The Call for Participation in MULSAIC'97 can be found at the WWW page:

http://www.iit.nrcps.ariadne-t.gr/~costass/mulsaic97.html