

MIETTA -- A Framework for Uniform and Multilingual Access to Structured Database and Web Information

Feiyu Xu, Klaus Netter, Holger Stenzhorn,

DFKI Language Technology Lab
Stuhlsatzenhausweg 3
66123 Saarbrücken, Germany
{feiyu, netter, holger}@dfki.de

Abstract

We describe a WWW-based information system called MIETTA, which allows uniform and multilingual access to heterogeneous data sources in the tourism domain. The design of the search engine is based on a new crosslingual framework. The framework integrates a crosslingual retrieval strategy with a strategy using natural language techniques: information extraction and multilingual generation. The combination of information extraction and multilingual generation enables on the one hand, multilingual presentation of the database content, and on the other hand, free text crosslingual information retrieval of the structured data entries. We will demonstrate that the new framework is useful for domain specific and multilingual applications.

Keywords: web documents, structured database information, multilingual and crosslingual information retrieval, information extraction and multilingual generation, machine translation

1 Introduction

In recent years, more and more attention has been paid to multilingual information access both in the crosslingual information retrieval ([2];[9];[15];[16]) and the natural language processing research communities ([1];[3]). Most approaches deal either with unstructured textual information like web documents or with structured information like relational database information. Systems which handle both kinds of information are rare. However, in real world applications, information providers in many domains, often have to provide access to heterogeneous data sources.

Our framework was developed within the project MIETTA (Multilingual Information Extraction for Tourism and Travel Assistance), a project in the Language Engineering Sector of the Telematics Application Program of the European Commission. The tourism domain is by its very nature multilingual ([18]), and tourist information is typically maintained as web documents or as database information by institutions like national or regional tourism offices. The main objective of MIETTA is to facilitate multilingual information access in a number of languages (English, Finnish, French, German, Italian) to the tourist

information provided by three different geographical regions: the German federal state of Saarland, the Finnish region around Turku and the Italian city of Rome.

In many applications, structured database information is accessed by means of forms, unstructured information through free text retrieval. In our approach, we attempt to overcome such correlations by making it completely transparent to the user whether they are searching in a database or a document collection, leaving it open to them what kind of query they formulate. Free text queries, form-based queries and their combination can yield documents and structured database information. The retrieved results are presented in a uniform textual representation in the user language.

We use automatic document translation to deal with web documents, because it allows the user to access the content without knowledge of the document language and provides good retrieval performance within our limited domain (see section 2). At the same time, multilingual access to the database information is supported by the combination of information extraction and multilingual generation. Information extraction extracts domain-relevant templates from databases and normalizes them in a language-independent format, while multilingual generation produces natural language descriptions from templates. As a result, the database content becomes multilingually available for the result presentation, and natural language descriptions can be handled in the same way as web documents, namely, we can apply the advanced free text retrieval methods to them.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and / or a fee.

Proceedings of the 5th International Workshop Information Retrieval with Asian Languages

Copyright ACM 1-58113-300-6/00/009 ... \$5.00

The challenge of the approach is to merge the technologies of crosslingual information retrieval and natural language processing to achieve the following goals:

- Provide full access to all information independent of the language in which the information was originally encoded and independent of the query language;
- Provide transparent natural language access to structured database information;
- Provide hybrid and flexible query options to enable users to obtain maximally precise information.

In the following sections we describe how these goals can be achieved in the MIETTA framework. The paper is organized as follows: Section 2 discusses the crosslingual information retrieval approaches and the natural language processing techniques which are related and used in the MIETTA framework. Section 3 describes the overall MIETTA system. Section 4 shows how much effort has to be paid to adapt the framework to new domains and new languages. Section 5 describes some evaluation issues for the MIETTA System. Section 6 summarizes the whole approach and discusses future work.

2 Framework Related Approach

In this section, we describe different options of crosslingual information retrieval, as well as the approach which combines information extraction and multilingual generation, and we discuss how these are applicable to the MIETTA context.

2.1 Crosslingual Information Retrieval

The main idea of crosslingual information retrieval is to search document collections in different languages by providing a query in one single language. Typical strategies are based on query translation, document translation or a combination of both. In the following, we will discuss advantages and disadvantages of the different strategies. Against this background, the technology decision in MIETTA can be easily understood.

2.1.1 Query Translation

The main goal of query translation is to help the user to formulate their query in another language, such that the translated query can then be used as a search term (see Figure 1).

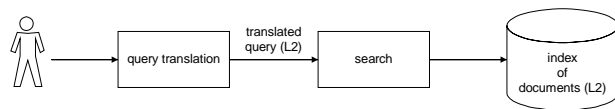


Figure 1: the workflow of crosslingual information retrieval based on query translation.

The primary problem of query translation is that short queries provide less context for word sense disambiguation,

and inaccurate translations can lead to bad recall and precision of the search results ([5]). The MULINEX system ([11]) provides a kind of user interaction for the sense disambiguation of translated terms. However, this kind of approach is only feasible in a specific scenario, namely, if the user has enough knowledge of the target language in order to select the right sense. Retranslating possible translations back into the original query language can solve this problem to a limited degree. Even then the crucial problem remains, that the user needs to have enough passive knowledge of the document language in order to understand the content.

2.1.2 Document Translation

In a document translation approach, the search index is built from automatically translated documents, such that the search becomes similar to a monolingual search, i.e., the user query can be used directly as the search term (see Figure 2).

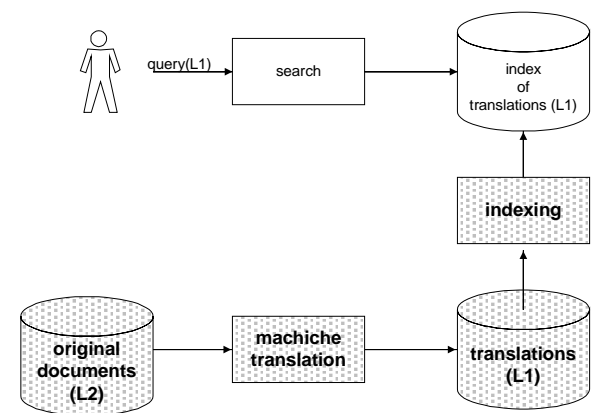


Figure 2: the workflow of the document translation based approach. Components with the gray background belong to the offline processing.

The advantage of this strategy can be a higher translation and retrieval accuracy, since the full original document provides more context for disambiguation. Although retrieval performance still heavily depends on the quality of the underlying MT system, the word sense disambiguation problem is less severe. Therefore, this option is often preferred compared to query translation ([5]; [10]). The main limitation is, of course, that under this approach at least the indices have to be duplicated, and in the offline translation the translated documents also need to be stored. Thus, for a universal or global search engine, this approach is practically not viable due to massive cost of computation and storage. However, the approach is quite suitable in a restricted domain where the number of documents is limited.

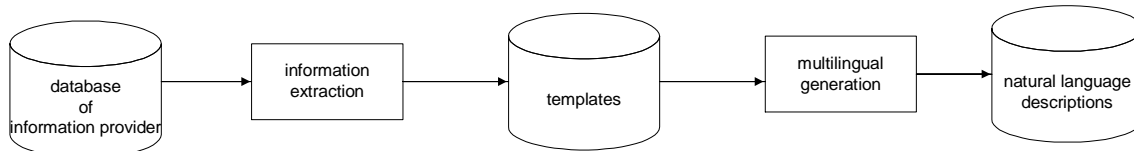


Figure 3: interaction of information extraction and multilingual generation

2.1.3 MIETTA Decision

In MIETTA, the document translation based strategy was preferred, as it allows for direct access to the content, and as it provides better performance within a restricted domain. The MIETTA web documents are limited to regional servers for the tourist domain, such that we do not face big storage problems. To translate the documents, the LOGOS¹ system was employed, which covers the following directions:

German ⇒ English, French, Italian

English ⇒ French, German, Italian, Spanish

The situation to start from in MIETTA was as follows: Rome could provide documents manually translated into English, French, German, Italian and Spanish; Turku had Finnish documents, most of which were also translated into English, while the documents in the Saarland were mainly in German. The final document collection in MIETTA after the document translation yielded an almost fully covered multilingual setup:

- Turku: Documents are available in all five languages. LOGOS has translated the manually translated English documents into three other languages, French, German and Italian.
- Rome: Documents were already available in all MIETTA supported languages except Finnish.
- Saarland: Documents exist in four languages except Finnish, after LOGOS translates the German documents into English, French and Italian.

2.2 Information Extraction and Multilingual Generation

The database information offered by the MIETTA information providers was mostly semi-structured and encoded in different languages (Italian, German and Finnish). Hence, most of the relevant information could only be found in the comment fields, mixed with other information. In order to make the database content more structured and multilingually accessible, we pursued an approach that combines information extraction and multilingual generation. The objective of information extraction in MIETTA is thus twofold:

- To extract the domain relevant information (templates) from the unstructured data so that the user can access more facts and more accurately;
- To normalize the extracted data in a language-independent format to facilitate the multilingual generation.

Multilingual generation takes a template provided by the IE component as input and generates a natural language description from it. The interaction of information extraction and multilingual generation is depicted in Figure 3.

A desired side effect of this strategy is that we can apply the same free text retrieval methods to the generated descriptions as to the web documents.

2.2.1 Information Extraction

The main task of information extraction is to analyze unstructured text and identify the relevant pieces of information ([8];[13]). One of its application areas is to detect the domain-relevant information from unstructured text and convert it into codified database entries. The relevance of the information is determined by templates, which are predefined for the domain. We describe in this subsection our usage of information extraction with the help of some examples.

The following German example text comes from the comment field in an event calendar from the Saarland:

(1)

St. Ingbert: -Sanfte Gymnastik- für Seniorinnen und Senioren, montags von 10 bis 11 Uhr im Clubraum, Kirchengasse 11.

(English: *St. Ingbert: -Gentle Gymnastics for seniors, every Monday from 10:00 to 11:00 am, in Club room, Kirchengasse 11*)

The above text contains three pieces of information about the event, namely, the name of the event, the location and the temporal duration of the event:

```
<event> <city> St. Ingbert </city>: <name> -Sanfte Gymnastik- für Seniorinnen und Senioren </name>, <time> montags von 10 bis 11 Uhr: </time> <location> im Clubraum, Kirchengasse 11 </location>. </event>
```

To extract the relevant pieces of information contained in texts like the above example, we designed three steps:

- *NL shallow processing*: Identifying the relevant chunks of the text, for example, noun phrases and named entities (date, time, location, geographic names, phone no. and addresses).

¹ See <http://www.logos-ca.com>.

- *Normalization*: Converting information into a language-independent format; for example, date, time, location, addresses and phone no.
- *Template Filling*: Mapping the extracted information into the database fields by employing specific template filler rules.

We applied the IE system SMES ([11]) for German shallow processing and IUTA² for Italian text analysis. The Finnish part is still to be extra developed. For the “normalization”, we defined uniform formats for date and time expressions, phone numbers, addresses, etc. For example, “montags von 10 bis 11 Uhr” (English: *Every Monday from 10:00 to 11:00*) is normalized as follows:

Start time:	10:00
End time:	11:00
Weekday:	1
Weekly:	Yes

For our tourist domain, a specific set of templates was defined, which corresponds to concepts like “event”, “accommodation”, “tours” etc. These concepts were organized in a three level concept hierarchy, drawing on the expertise of the MIETTA user partners in the tourism sector. Even if the hierarchy was designed for the MIETTA users, it can be easily adapted and generalized to other regions of tourism interest. The underlying format of the concept hierarchy is language independent. Each general concept can have several daughters, for example, “event” has “theatre”, “exhibition”, “cinema” and “sports” as its sub-events. Templates in the same concept hierarchy inherit all attributes from their parents. For example, all the event templates have location, time and title as their attributes.

2.2.2 Multilingual Generation

In recent years, shallow natural language generation approaches have been shown to be quite useful for realistic applications within limited domains. In particular, the combination of information extraction and multilingual generation provides a useful approach to a multilingual information presentation of structured information into a textual format. The basis for shallow text generation applied in MIETTA is the system TG/2. This system was developed in the TEMSIS project ([4]), whose objective was to generate summaries of environmental data in German and French from database information. We use a JAVA implementation of TG/2, called JTG/2³. JTG/2 takes some language-independent input, applies language specific grammar rules and morphological lexicon, and returns some language-specific description. In MIETTA,

² <http://celi.sns.it/~celi/projects/Iuta/iuta-top.html>

³ JTG/2 is a Java implementation of DFKI's TG/2, developed by CELI, the Centro per l'Elaborazione del Linguaggio e dell'Informazione, for more information see <http://www.celi.it>.

five language-specific grammars were developed for the template generation. Because the JTG/2 rule formalism supports shallow grammar rules, construction new language-specific rule sets requires comparatively little effort. We illustrate the approach through a simple example from a MIETTA template. An event, such as a theater play, is encoded in a corresponding template as follows:

(2)

Level1:	Event
Level2:	Theater
Level3:	_
Event-Name:	Faust
StartDate:	21.10.99
PlaceName:	Staatstheater
Address:	Schillerplatz 1, 66111 Saarbrücken
Phone:	06 81-32204

The above template is used as the input for generation into five languages, resulting for example in the following texts:

(3)

English:

The theater show Faust will take place at the Staatstheater in Schillerplatz 1, 66111 Saarbrücken (in the downtown area). The scheduled date is Thursday, October 21, 1999. Phone: 06 81-32204

(4)

Finnish:

Teatteriesitys Faust järjestetään Staatstheaterissa, osoitteessa Schillerplatz 1, 66111 Saarbrücken (keskustan alueella). Tapahtuman päivämäärä on 21. lokakuuta 1999. Puhelin: 06 81-32204.

Texts like above are employed in two forms in MIETTA: they are used on the one hand as result presentation of the template content, and on the other hand, as input for free text indexing to allow advanced free text retrieval.

2.2.3 Conclusion

The combination of information extraction and multilingual generation has several advantages: It can make the translation of textual information in a database unnecessary, thus saving the duplication of the same piece of information in different languages. One important advantage is, of course, that it greatly facilitates the maintenance of such data and it ensures a higher degree of consistency across different languages. At the same time, it allows for an integrated or hybrid free text retrieval approach to both structured database and document information with the added dimension of multilinguality.

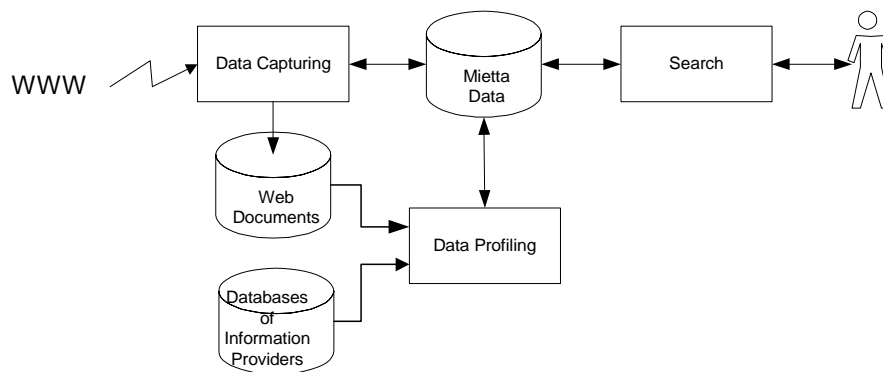


Figure 4: Miatta Data contains all data that is produced and processed, and that is accessible throughout the system. It contains different kinds of data, including for example, translations, templates, texts generated from templates, free text indexes etc. It contains the output of the Data Capturing and the Data Profiling components and the input for the Search Engine.

3 The MIETTA System

The user requirements of a tourism information system are quite varied with respect to the content. The needs for information range, on the one hand, from fairly structured information and precise facts (such as the opening times of a museum or the price of hotel accommodation), to some more general background information, as it is typically described on web pages concerning certain regions, towns, or vacation facilities. While the former type of information will be typically stored in a structured format (as a relational database), the latter is mostly available only in an unstructured format (as text documents). In order to allow the user to access these two different sorts of information in a uniform way, we provided hybrid search options:

- *Free text retrieval*: The user can enter several words or phrases to find both web documents and descriptions generated from templates.
- *Concept based navigation*: The user can navigate through web documents and templates according to the MIETTA concept hierarchy.
- *Form-based search*: The user can select fields in a search form to access templates.

Our motivation is to make it completely transparent to user which source of information they are searching in, and to allow them to formulate their query as precise as they desired. In order to realize our goal, we developed our framework which integrated intelligently the existing techniques occurred in the crosslingual free text retrieval and natural language processing communities. The implementation of the framework is our MIETTA system, which contains the following three main components:

- *Data Capturing*: The task of this component is to account for the collection of web documents and for recording document information such as the title, URL, manual and automatic classifications, etc.
- *Data Profiling*: This component comprises several components primarily related to our framework, such as document translation, information extraction of database information, offline multilingual

generation of templates, offline free text indexing of web documents, translations and generated descriptions.

- *Search Engine*: It enables the end user to search and access the search result.

The interaction of the three components is illustrated in figure 4.

3.1 Data Capturing

A large part of the basic material provided by the MIETTA user partners comes in the form of web documents. To make this information accessible through the MIETTA search tools, they must be registered, gathered and indexed. The data capturing tool provides a convenient user interface which allows the information providers to register their web pages. The registration user interface provides a broad range of facilities: The user can enter the URL of the page or the site, the depth to be disclosed, and to enter the address of a potential document translation. Furthermore, he can also classify the page content based on the MIETTA concept hierarchy. Besides the URL registration and manual classification, the data capturing tool also integrates the JumboScan Web crawler package provided by UNIDATA⁴ for downloading and indexing the document information like the URL name, title etc.

3.2 Data Profiling

The aim of the data profiling component in the MIETTA system is to disclose the data sources in such a way that access through different search options becomes possible. It contains

- Document translation, based on the LOGOS machine translation system;
- Information extraction from database entries for template construction;

⁴ See <http://www.unidata.it/>.

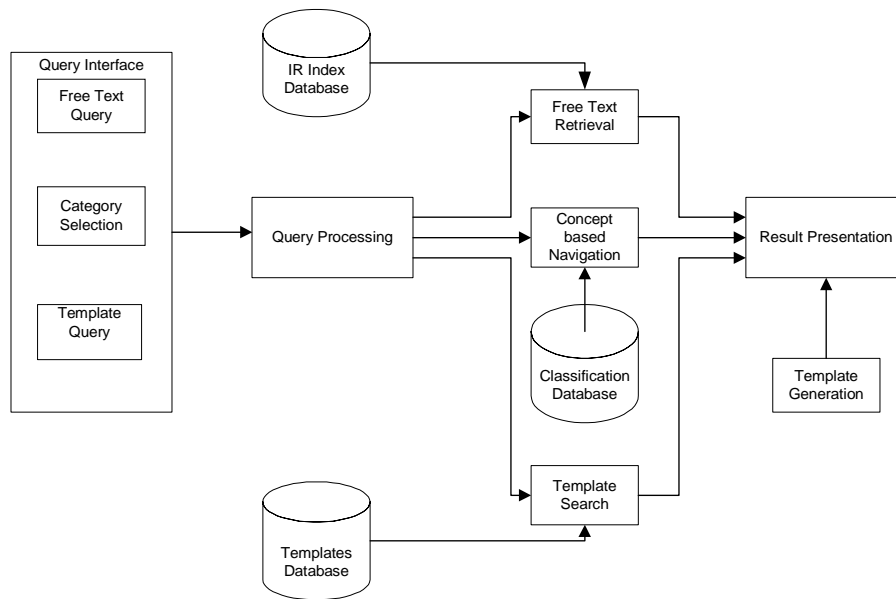


Figure 6: the architecture of the search engine

- Multilingual generation from templates to obtain natural language descriptions;
- Free text indexing

The first three aspects have been discussed in the last section. The result of applying these processes is that both web documents and template information becomes available in the different languages covered by MIETTA. We will focus on the free indexing work here. We employ an existing indexing and search tool developed by TNO⁵ ([12]) for the free text retrieval task. The TNO indexing tool generates two kinds of indexes:

- A lemma-based fuzzy index based on trigrams (ISM);
- A Vector Space Model (VSM) index based on lemmatas.

These free text indexing components are applied to web documents, translations and descriptions generated from templates, see the following Figure 5.

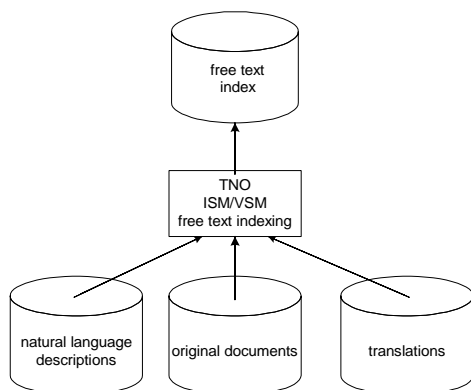


Figure 5: free text indexing

The indexing of the automatically generated descriptions adds a specific functionality to the MIETTA search engine, as the template content becomes accessible for fuzzy search and the vector space model search, which are normally not supported by relational databases. As a result, both web documents and database information becomes available in a textual format in different languages, disclosed through classifications and a free text index.

3.3 Search Engine

In this subsection, we describe the hybrid search options provided by MIETTA and how they are realized. With the help of data capturing and data profiling, the MIETTA search engine allows for standard free text retrieval as well as the following advanced search capabilities:

- Concept-based navigation
- Form-based query (template search)

The architecture of the search engine is described in Figure 6.

As mentioned above, MIETTA uses the existing TNO ISM/VSM search engine for free text retrieval. The ISM part makes use of a kind of a fuzzy matching algorithm based on trigrams. It allows to match index terms with query words or phrases containing spelling errors or morphological variants. For example, the user can enter “baroque palaces” and find documents and template descriptions which contain the phrase “baroque styled palace”. In addition to the free text retrieval, the user can also navigate through the concept hierarchy to search for information in a certain category. In contrast to many other search engines, the MIETTA user can also combine the free text retrieval with the concept-based navigation by formulating a query with constrains such as “find all

⁵ See <http://twentyone.tpd.tno.nl/>.

Figure 7: search menu

Figure 8: form based query

documents containing the word *colosseo* belonging to the category *Art and Culture*”, see Figure 7.

A more restricted and goal-directed query is the form-based query, where the user can select fields in a template form. For example, the user can select the “time” and the “location” fields of a “concert” event template by using a query form. In Figure 8, the user has formulated a query corresponding to the constraint “give me all information about concerts in the city center today”.

All queries are processed by the query processing component and converted either into a standard SQL query or an ISM/VSM query. The result of the retrieval is presented as a uniform list of links to textual descriptions (generated from templates) and to web documents. Both types of information are presented, on the one hand, in an absolute ranking order, where only the relevance of the document plays a role, and on the other hand, sorted according to the different categories. If the user clicks on a link, they receive either a web document or a generated text as in the examples (3) and (4).

To summarize, the MIETTA search engine represents a flexible way of combining crosslingual free text retrieval with standard database access. The hybrid query options and their interaction provide the user with a highly versatile range of options to express their different search requirements, which is also reflected in the presentation of the results and the further navigation options.

4 Scalability of the Framework

Our framework can be easily adapted to other domains. The domain modeling consists mainly of the definition of domain specific templates and the concept hierarchy. With respect to the information extraction task, the major part of the effort needs to be spent on the new definition of the template filler rules, since the natural processing and the normalization steps are domain independent. The MIETTA multilingual generation tool has already been proven to be reusable, as it has been applied in both the TEMSIS environmental and the MIETTA tourist domains.

When it comes to adding a new language, three language dependent components would be involved in the

MIETTA system: natural language generation (JTG/2), document translation and natural language processing. As mentioned in section 2, our natural language generation tool requires less effort for the development of a grammar rule set in a new language. It supports also easy integration of a morphological component. In addition to the five generation grammars developed within MIETTA, we also carried out some successful experiments with a Chinese grammar. Thus, integrating a new language outside the Western language family into JTG/2 also appears to be quite easy. Document translation is clearly dependent on the language pairs supported by the machine translation system employed, i.e., essentially independently from the MIETTA system itself. However, to deal with the problem of unavailable translation pairs, a statistical translation model, as proposed in ([6]) could be employed. Such translation models are much easier to establish than MT systems, and it would be sensitive to the domain of the training corpus. Similar to the scalability of the machine translation, the shallow language processing components are dependent on the developments for a certain language. In the recent years, the natural language processing society in Asia is very active ([7]), in particular, as far as Chinese, Japanese and Korean are concerned. Hence, the integration of one of these languages is realistic.

5 Evaluating MIETTA

Because of the board variety of search strategies and the heterogenous data sources, the standard relevance assessment model used in the ad hoc and routing forums of TREC is difficult to apply to the complete MIETTA system. The evaluation of the individual components such as the TNO free text retrieval engine, the natural generation system and the IE tools SMES and IUTA can be found in their corresponding literature mentioned in the previous sections.

The strategy, we plan to explore, is an end user centered evaluation. We will set up a questionnaire which contains a list of MIETTA specific queries and different evaluation criteria concerning

- The transparency and ergonomics of the search user interface and the result presentation,
- The quality of machine translation and natural language generation,
- The interaction of the hybrid search options.

6 Conclusion

We have presented a novel framework for the uniform and multilingual access to web documents and the structured data and have implemented a practical application that successfully realized this framework. The MIETTA system allows the user to carry out a crosslingual search in different sources of information at different levels of content granularity. This framework is highly suitable as a domain-specific information system and internet portal. It can be easily transferred to other domains and is extensible to other languages.

Future work will be directed towards extending the framework to information extraction from web documents and to fully automatic document classification. In the current MIETTA system, template extraction from web documents combined with multilingual generation has not been considered due to limited resources. Such a combination, however, would make the system even more effective as it could provide both summaries of web documents and multilingual access to such summaries.

Acknowledgements

This approach is grounded on the cooperation in the MIETTA consortium. The MIETTA consortium consists of the following institutions: The technical partners are DFKI (Deutsches Forschungszentrum für künstliche Intelligenz), CELI (Centro per l'Elaborazione del Linguaggio e dell'Informazione), Unidata S.p.A., the University of Helsinki and Politecnico di Torino. The user partners are the city of Rome, Staatskanzlei des Saarlandes and city of Turku and Turku TourRing. More information can be found in <http://www.mietta.net>.

We are grateful to Paul Buitelaar, Olga Goldmann and to our colleagues at the MIETTA Partner institutions: Luca Dini, Vittorio Di Tomaso and Giampaolo Mazzini in CELI, Alessandro Giarante, Marcello Vispi in Unidata S.p.A., Kimmo Koskenniemi, Jyrki Niemi in the University of Helsinki, Elena Baralis and Rosa Meo in Politecnico di Torino, our project coordinators Francesco Bellini and Audrey Boss in Comune di Roma and our user partners.

References

- [1] Appelt and Israel: Building Information Extraction Systems. *ANLP-97 Tutorial*, 1997.
- [2] Braschler, Croal Peters and Peter Schäuble: Cross-Language Information Retrieval (CLIR) Track Overview. In *Proceedings of the eighth Text REtrieval Conference (TREC-8)*, held in Gaithersburg, Maryland, November 17-19, 1999.
- [3] Stephan Busemann: Language Technology for Transnational Web Services. In *Proceedings European Telematics: Advancing the Information Society Telematics Applications Programme Annual Concertation Meeting*, Barcelona, 1998, pp. 101-105.
- [4] Stephan Busemann and Helmut Horacek: A Flexible Shallow Approach to Text Generation. In *Proceedings of the Ninth International Natural Language Generation Workshop (INLG '98)*, Niagara-on-the-Lake, Canada, August 1998, pp. 238-247.
- [5] Jaime Carbonell, Yimying Yang, Robert Frederking, Ralf D. Brown, Yibing Geng and Danny Lee: Translingual Information Retrieval: A comparative evaluation. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, August 1997.
- [6] Jiang Chen and Jian-Yun Nie: Parallel Web Text Mining for Cross-Language IR. In *Proceedings of 6th International Conference on Computer-Assisted Information Retrieval (RIAO 2000)*, Paris, 2000.
- [7] "Closing the Millenium". In *Proceedings 5th Natural Language Processing Pacific Rim Symposium*, Beijing, China, 1999.
- [8] Cowie and Lehnert: Information Extraction. In *Communications of ACM*, 39(1):51-78, 1996.
- [9] Mark W. Davis and William C. Ogden: Implementing cross-language text retrieval systems for large-scale text collections and the world wide web. In *AAAI Symposium on Cross-Language Text and Speech Retrieval. American Association for Artificial Intelligence*, March 1997.
- [10] S. Dumais, T. Landauer and M. Littman: Automatic Cross-Linguistic Information Retrieval using Latent Semantic Indexing. In *Proceedings of SIGIR-96*, Zurich, August 1996.
- [11] G. Erbach, G. Neumann and H. Uszkoreit: MULINDEX - Multilingual Indexing, Editing and Navigation Extensions for the World Wide Web In *David Hull and Doug Oard (eds.) Cross-Language Text and Speech Retrieval -- Papers from the 1997 AAAI Spring Symposium*, AAAI Press, Menlo Park, 1997.
- [12] Djoerd Hiemstra and Wessel Kraaij: Twenty-One in ad-hoc and CLIR. In *Proceedings of the Seventh Text Retrieval Conference (TREC-7)*, E.M. Voorhees and D. K. Harman (editors), NIST special publication, 1998, pp. 500-240.
- [13] R. Grishman and B. Sundheim: Message Understanding Conference - 6: A Brief History. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, Kopenhagen, Denmark, Europe, 1996, pp. 466-471.
- [14] K. L. Kwok: English-Chinese Cross-Language Retrieval based on a Translation Package. In *Workshop on Machine Translation for Cross Language IR - MT summit VII.*, Signapore. 1999, pp. 8-14.
- [15] Douglas W. Oard: Global Access to Multilingual Information. Presented at IRAL99, Taipei, 1999.
- [16] Douglas W. Oard: Cross-Language Text Retrieval Research in USA. In *3rd ERCIM DELOS Workshop*, Zurich, Switzerland, 1999.
- [17] J. Piskorski and G. Neumann: An Intelligent Text Extraction and Navigation System In *Proceedings of 6th International Conference on Computer-Assisted Information Retrieval (RIAO-2000)*, Paris, 2000.
- [18] Tschanz and S. Klein: Web-enabled Cooperation in Tourism. In *Proceedings of EMOT Workshop*, Modena, Italian 1996.