Cross-Lingual Information Transformation and Summarisation for Portable End-Devices with the MEMPHIS Toolbox

Walter Kasper
DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
Kasper@dfki.de

Katarina Stanoevska-Slabeva
mcm institute
University of St Gallen
Blumenbergplatz 9
CH-9000 St. Gallen, Suisse
Katarina.Stanoevska@unisg.ch

Yiannis Stavroulas
Angelos Yannopoulos
National Technical University of Athens
Institute of Communications and Computer Systems
9, Heroon Polytechneiou Street
EL-15773 Athens, Greece
ystavroulas@telecom.ntua.gr
ang@telecom.ntua.gr

Abstract

The MEMPHIS project (http://www.ist-memphis.org) aims at developing a platform for cross-lingual premium content services, targeting mainly portable thin clients, like mobile phones, PDAs, etc. The core of the system is a cross-lingual transformation layer integrating cross-lingual information extraction and summarization of source documents, translation to the customers' target languages as well as multi-lingual generation of documents according to the restrictions and requirements of the various target devices for distribution.

1 Introduction

The possibility to access information available on Internet through mobile devices anywhere and anytime has given rise to high expectations for providers and users of information. In particular providers of information expect to be able to deliver information to users and customers in a more personalised way. But, mobile access to Internet sources and delivery of information to mobile devices is in many cases a cumbersome process because of the following obstacles:

- Due to the small screen size of portable devices, it is almost impossible to view available information online in an efficient manner without dedicated transformation of information for mobile devices.
- Usage patterns of mobile users show, that mobile devices are used in many cases during niche times for example during waiting for and using public transportation [1], [9]. Most of the time, there is no specially dedicated time for mobile browsing as in the case of Internet. As a result information has to be provided in an efficient manner according to the specific personal needs of user.
- The mobility of users results in many cases in the need to access local information for example during holidays or business travels. But, in many cases the information might be available in a local language, which might be different than the languages understood by the user.

Several approaches can be observed in literature and in practice how to cope with the above mentioned problems: One solutions would be to develop special Web pages for small devices by using the wireless markup language (WML). While this is a very effective approach for new information, it requires considerable transformation effort to adapt existing information. Automatic transformation procedures as proposed by [5] support real-time transformation of available information, but in many cases not all information elements can be meaningfully transformed.

Another approach is the automatic miniaturisation of standard HTML page as for example proposed by [2]. A drawback of such a system is a need for frequent scrolling in order to assess all relevant information elements. A third approach is to apply natural language processing techniques and automatic summarisation (see for example [3]) in order to meaningfully reduce the size of available information and adapt its representation to different end devices.

Most of the mentioned approaches are dedicated to solutions supporting mobile users during browsing and access to available online information and do not provide multi-lingual support. Another possible approach is to assess or collect information needs of mobile users in advance and to provide them with relevant information in a push manner. This allows for a more efficient transformation processes as relevant information can be transformed once and delivered to many users with similar needs. But, it limits the flexibility of the user to deal with ad-hoc information needs.

In this paper we will present the MEMPHIS project and toolkit, which extends existing approaches for automatic summarisation of online information with a multi-languge transformation layer for mobile information services. Based on Natural Language Processing Technologies (NLP) for translation and summarisation of information, the MEMPHIS project (http://www.ist-memphis.org) aims at developing a platform for cross-lingual premium content services, targeting mainly portable thin clients, like mobile phones, PDAs etc. Users of MEMPHIS register at the MEMPHIS portal where they can choose among certain topics of interest, as well as in which target language, in what form, on what device and how often they want to be informed. The MEMPHIS system undertakes to find relevant content in the WWW, transform and translate it according to each individual user's preferences and finally deliver it using the user's preferred delivery method and device. Major project result is to be a set of NLP modules and a toolkit which allows easy adaptation of the modules to new domains in order to set up new services.

As proof of concept, two applications are developed and tested within the project, which differ considerably with respect to their domains, the information sources and service requirements:

- *MediAlert*, a service that provides information about new books on specific topics, e.g. music or biographies.
- *FinAlert*, a service that provides information about mergers and acquisition activities in the banking area.

Prototype systems for both applications are available and are currently tested by external users. This paper will present an overview of the MEMPHIS toolbox and first results of internal tests. The content of the paper is structured as follows: Section 2 presents the overall architecture of the MEMPHIS system. Section 3 describes the components of the MEMPHIS toolbox and the information transformation steps in more detail. Section 4 presents the two services and a summary of results from internal tests. Section 6 closes the paper with an outlook.

2 Overall Architecture of the MEMPHIS System

Figure 1 presents an overview of the main modules of the MEMPHIS system and their functionality.

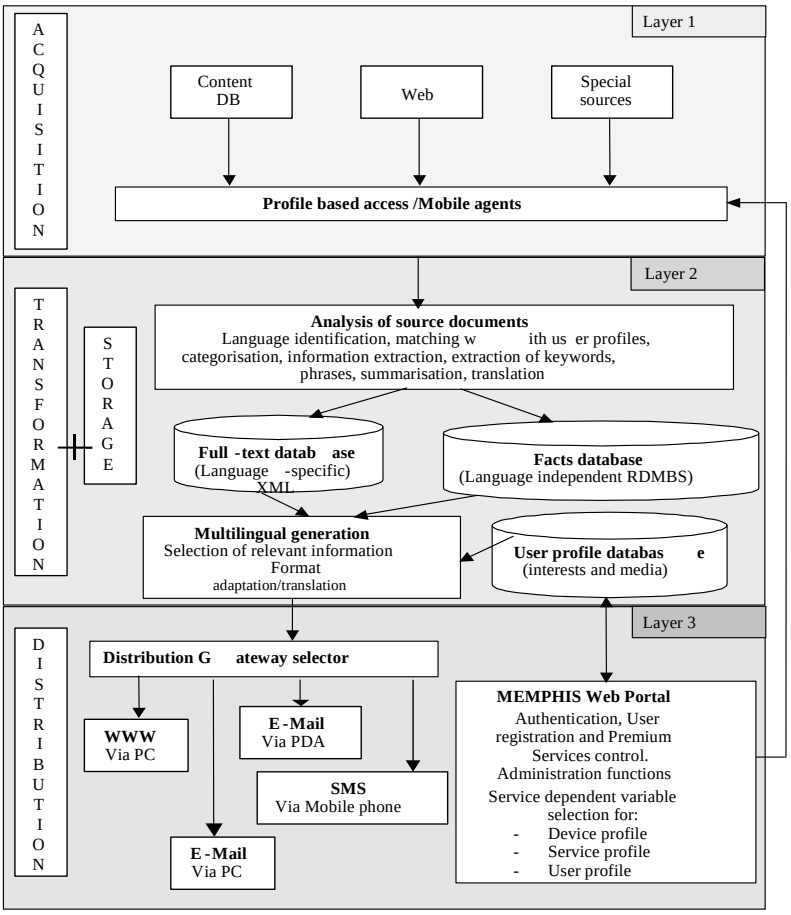

Fig. 1: **MEMPHIS system block scheme**

The modules are grouped according to their functionality in three layers: 1) the acquisition layer, 2) the multi-language transformation layer, and 3) the distribution layer (c.f. 1). The three layers denote the sequence of document processing in MEMPHIS.

The *acquisition* layer retrieves documents from various sources according to the user and service profiles. The *transformation* layer analyses and extracts relevant information from the acquired document and stores it as basically language independent "factual" relations using the domain ontology. Multilingual generation then is employed to generate the textual presentation of these facts. Machine translation supports the generation with respect to language dependent factual content. The *distribution* layer finally sends the documents to the devices the user specified.

The user profiles are defined and managed through a web portal allowing the user to specify his devices as well as selection of both delivery parameters and of specific topics of interest from those the service offers. Delivery parameters include the device the information should be sent to, the time of delivery - e.g. at once, once a day, once a week, etc. - the preferred language for the information presentation and summary size. The available information is classified in main categories as for example financial

or media information services. In order to provide the best possible quality of translation with state-of-the-art technology each main information category is further divided in special topics. For example specific topics of financial information services might be “Bank merger and acquisition”, “Stock exchange information” and similar.

In a runtime version of the system the data exchange between the different modules is co-ordinated by two databases the user and device database as well as the document database. The synchronisation of the modules is controlled by the MEMPHIS monitor, which controls all the system components to guarantee proper system behaviour (c.f. 2). The monitor also contains a database handler which is in charge of the system’s databases and supports the three processing layers by providing a database access API.

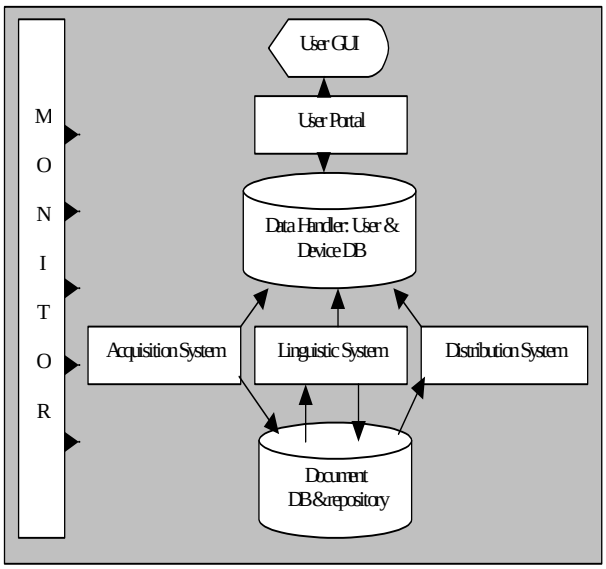

Figure 2: MEMPHIS Runtime System Overview

In the following, we will describe in more detail the information transformation in the system. We think, the special merits in MEMPHIS lie in the integration of many of current language technologies and approaches in one coherent system to achieve an innovative application functionality.

3 Components of the MEMPHIS System

3.1 Content Acquisition

The Acquisition subsystem of MEMPHIS is dedicated to finding and fetching content from the Internet to the MEMPHIS home platform. It does not perform web crawling; only known, pre-determined content providers’ sites are used, so that the credibility of the content the system delivers can always be guaranteed. Content classification and all similar high-level processing of the documents is performed by the cross-lingual transformation subsystem, however the acquisition subsystem performs a first filtering of the content. This serves both as a bandwidth-saver, since it would be very bandwidth-intensive to draw all the textual content found in a content provider’s site

to the MEMPHIS platform, and as a pre-processing step to the linguistic processing, drastically cutting down on the workload of the linguistic transformation subsystem.

The key requirements of acquisition are efficiency, flexibility and security. The whole subsystem is based upon Mobile Agent technology and the mobile agent platform Grasshopers is employed. In order to be efficient, the Acquisition system employs a distributed search architecture. Specifically, an agent platform can be set up and configured at the premises of each content source, so that searches can be executed locally on the content source's site. Thus minimal bandwidth is spent in searching the Internet, as only useful documents are returned to the main system.

Flexibility in the configuration and deployment of the system is considered to be of paramount importance. Distributing the system by setting up search platforms in the content providers' premises is a powerful capability, however it can be very restrictive, as providers may not be willing to host a third-party system at their sites, even at the cost of forsaking the business benefits that a sophisticated search capability could trigger. Mobile agent technology waives these considerations easily, as the agents executing the search are insensitive to the search site's location. The agents use a directory that maps content providers' sites to agent platforms. This way, the system can employ virtually any configuration transparently. For performance reasons, it is always better to set up a search site in the content provider's premises but, if this is not possible, any other platform – including the MEMPHIS home platform – can be designated as the search platform for this particular content provider. Hence an agent wishing to execute a search at this content provider's site will simply retrieve the destination platform from the directory and migrate to it, if not already there, regardless of the actual location of the platform. A provider-specific wrapper is used on the search site to manage the translation of data from the site-specific format to a MEMPHIS-standard format, understood by the agent.

Flexibility has another aspect in the MEMPHIS acquisition subsystem, that of the filters used. Since the acquisition subsystem may be called upon to search and fetch content from virtually any site, the filtering method must be highly configurable. Also, many sites have built-in search engines; being able to employ them is very effective in many cases. To achieve maximal adaptability of the filtering mechanism, we have created a framework for the transparent application of filters, as far as the search agent is concerned. This framework defines a standard interface for the filter, as well as between the agent and the content source wrapper. The latter is used by the search agent to request data from the source wrapper. The search agent submits the filter as part of this request. Subsequently, the source wrapper retrieves data from the source, transforms them to the acquisition standard format, and uses the filter's interface to pass them through it. Documents that successfully pass the filter are returned to the agent and eventually carried back to the MEMPHIS home site.

Multiple search algorithms can be used with this framework, as no component has any knowledge about a filter's specifics. We have thus planned for the use of Natural Language Processing filters, such as topic identification, to be loaded onto the mobile search agents and used for the execution of distributed searches. Furthermore, this framework makes it possible to use site-specific search facilities. This is achieved as follows: a site-specific filter is defined, which is 'dummy' in that it contains no search functionality, but only serves for passing parameters to the source wrapper. When a content request is passed to the source wrapper by a search agent, the wrapper identifies the filter type and triggers the appropriate site-specific search mechanism (i.e. by filling in an HTML form and submitting it). The results are processed to adapt their format and they are returned back to the search agent, which is completely unaware of all this process.

Last but not least, the acquisition subsystem can be used to power full-text searches of the providers' content, even when this is premium content and as such,

non-disclosable without a fee [8]. The key idea here is that, since the searches are executed locally on the provider's site, the premium content never leaves the provider's premises. The source wrapper can be configured in such a way, that although the full text of documents is given to the filter for evaluation, only the publicly disclosable parts of them (such as the title and abstract) are returned to the search agent. This capability is not exploited by the present applications of the MEMPHIS system, nevertheless it constitutes a major design strength which adds value to the MEMPHIS toolkit.

After completing a search, the mobile search agent eventually migrates back to the MEMPHIS platform. The fetched documents are stored in the document repository and follow the processing chain described in the following section.

3.2 Cross-Lingual Transformation Technologies

In order to be independent of specific document encodings like HTML, RTF, PDF or some XML based format, all documents are first converted to a standard XML format, the *Memphis File Format (MFF)*. This makes the linguistic systems independent from the specific format of the source document. A MFF document consists of a header section containing meta-data about the document and a body section with the content of the document. Meta-data consist of information like the language of the document, its source, its topics, etc. In the body the document is represented in a canonical form, abstracting from the format of the source document. This canonical form maintains besides the textual content only certain important properties of the source document, such as its paragraph structure, headers and titles as well as all kinds of font changes, including emphasized parts of the documents. At present, there are converters for HTML, NewsML and plain text into the MFF format. The MFF files are stored in the document database as source documents.

The documents are than processed by the modules of the transformation layer. The basic components of the Transformation Layer in Memphis are (c.f. 3):

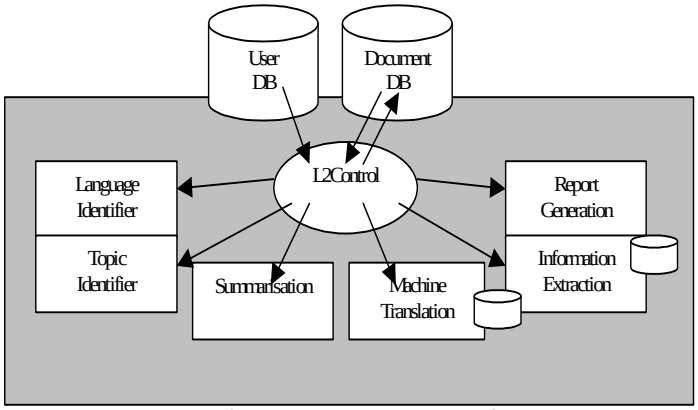

Figure 3: The Linguistic System Architecture

- Language identification and verification
- Shallow analysis: Tokenization, Segmentation, Morphology
- Named Entity Recognition
- Summarization analysis and Keyword extraction
- Application and service specific information extraction
- Machine Translation

- Multilingual generation

All levels of processing communicate through XML annotations on the source documents. These annotated documents can be stored as analysis documents in the database for future use to reduce the necessity to repeat all analysis steps, e.g. for indexing in a cross-lingual retrieval scenario.

3.2.1 Language identification

As a prerequisite to all linguistic processing the language, the source document is in, must be known. Since MEMPHIS retrieves documents in various languages and the documents usually do not specify the language, the first step consists in identifying the document language or verifying it in case it is already given. For this purpose a robust statistical approach based on Markov models is used. This approach performed best in a test on a corpus of 25000 documents from the MEMPHIS domains against a rank difference approach (TextCat, Cavnar(1993)) and a KNN based approach. Especially, the MediAlert domain of music presents difficulties to language identifiers as e.g. German documents in that area often contain foreign language terms.

3.2.2 Shallow linguistic analysis

The basis for all subsequent processing steps is a shallow analysis of the source document which tokenizes the documents and segments it into text units, e.g. sentences. Tokens are further subjected to morphological analysis. The morphological analysis provides stems used as normalized terms and part-of-speech information. So shallow analysis thus adds two levels of annotation to the source document as illustrated below in Figure 4. The `<p>` tag in the body denotes a paragraph. The segmentation adds the `<tu>` tags for textual units like sentences, the tokenization adds the `<Token>` tags while the morphology adds to the token the `<Reading>` tags for the possible analyses. Tokenization and segmentation are controlled by language specific parameter sets for classifying tokens. The token class is represented by its `type` attribute. All upper levels of analysis use these annotation levels.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE MFF SYSTEM "SummarizerFF.dtd">
<MFF version="1.0">
  <header>
    <language>de</language>
    <length>14702</length>
  </header>
  <body>
    <p>
      <tu>
        <Token string="Die" type="TOK" offset="1" length="3">
          <Reading stem="die" pos="Det" />
          <Reading stem="die" pos="Pron" />
        </Token>
        <Token string="beginnende" type="TOK" offset="5" length="10">
          <Reading stem="beginnen" pos="Adjective" />
        </Token>
        <Token string="Rezession" type="TOK" offset="16" length="9">
          <Reading stem="Rezession" pos="Noun" />
        </Token>
      </tu>
    </p>
  </body>
</MFF>
```

Figure 4: Analytic Document XML Markup

3.2.3 Named entity recognition

A further level of analytic annotation is added by a named entity recognition system. The system identifies person names, company names, location terms and date expressions. The current system is a pure statistical n-gram based tagging system trained on a large corpus of business news. It will be combined with a rule based system (SPROUT) in the near future, as it is difficult to find suitable and sufficiently large corpora for adapting the statistical system to specific requirements and other classes of named entities. The identification especially of person names and company names is important because these often consist of otherwise "unknown" words. So other later modules profit from this analysis by getting information on these items. Main users of this annotation level are information extraction and machine translation. Machine translation is improved considerably when the system can identify names as items not to be translated (cf. Eisele/Kasper(under review)).[1]

3.2.4 Summarisation

For summarisation a robust multi-lingual system following an extraction approach was developed which combines several types of heuristics: statistical term weighting in the *tf.idf* tradition, positional information and layout characteristics unit a relevance score. This assignment is used in generating indicative summaries of the source document of different sizes. One feature of the summariser is its query-adaptivity, that is it can be made sensitive to give user provided query terms special attention. This feature can be employed to generate different personalised summaries from the same text according to different user interests and topics.

Keyword extraction is an additional functionality using the same statistical models as the summarisation process. Therefore it was integrated closely with the summarisation process as an optional side effect of summarisation analysis.

3.2.5 Information Extraction

Information Extraction is the central functionality of the MEMPHIS system as it extracts from the documents retrieved that essential pieces of information the user is interested in. Basically this defines the MEMPHIS service as information service. This extracted information is stored as a factual relation using the domain ontologies to map the information from the documents to source document and language independent domain concepts. In practise, this relation is specified as an XML-DTD defining a template the extraction should instantiate. For MediAlert, the relevant information is information about books. Figure 5 illustrates an instantiated relation which is stored in the database as processing result for distribution to subscribed users. Several techniques are used for information extraction based on the type of documents to process:

Documents consisting of free text as for example offered by online news sites require a semantic based analysis of the text based on the domain ontologies. Corpus based terminology extraction is used to identify relevant terms related to specific pieces of information and concepts. These terms are used as triggers in a cascaded finite state extraction system. Not always the extraction succeeds.

Strongly structured documents as in the case of the MediAlert publisher announcements allow to exploit the structure to get at the information in more direct and reliable ways by defining a description of the document structure to locate relevant information. The drawback of such a system sensitive to specific document formats is

[1] In MEMPHIS, the Logos Machine Translation system from GlobalWords, Eisenach is used. All large scale MT systems, we are aware of, have problems with proper names.

that it is not very robust: when the publisher changes his presentation, the document descriptions need be updated accordingly. Nevertheless, we found that such an accommodation to a new presentation often does not take much effort. Even in this case a more generic linguistic based system must be kept in the background for less reliably or clearly structured document types, but also to control whether the structure driven system produces valid instances.

```
<BOOK doc-lang="en">
  <TITLE>DAY IN THE LIFE, A</TITLE>
  <AUTHOR>Mark Hertsgaard</AUTHOR>
  <ISBN>0-385-31517-1</ISBN>
  <PUBLISHER>Delta</PUBLISHER>
  <RESELLER>Random House</RESELLER>
  <YEAR>1996</YEAR>
  <PRICE_ORIGINAL>$ 15.95</PRICE_ORIGINAL>
  <SUMMARY_ANALYSIS>
     <tu id="9" weight="10.416451638392827" rank="1"> … </tu>
  </SUMMARY_ANALYSIS>
</BOOK>
```

Fig 5: MediAlert Information Extraction

3.2.5 Machine Translation

Translation is used for translating extracted relational information for multilingual output generation. Text translation is needed for summaries. Since fully automatic high quality translation at the present state of the art is only possible within very narrow domains and applications, the goals in MEMPHIS are more modest: similar as the extraction based summarisation, translation can only be an indicative abstract in the sense that it shows the user whether the document is relevant for him, so that he might look at the original, the translation quality should only indicate to the recipient to a sufficient degree that the information he got is indeed of interest to him.

When Memphis extends its functionality to provide for pull services in addition to the current push services, the translation will also be used for query translation as well as in the acquisition and analysis phase for cross-lingual indexing of documents (cf the results in the CLEF activities (cf. http://clef.iei.pi.cnr.it)).

3.3 Distribution

The distribution consists of four major modules (c.f. 4):

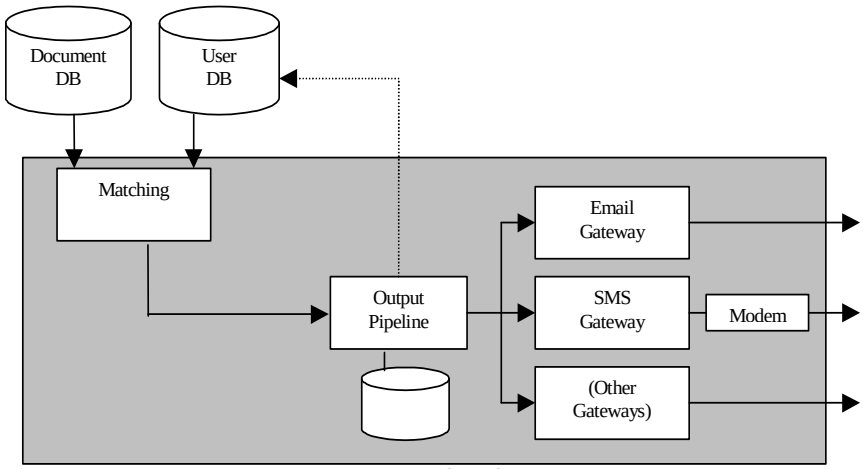

Fig. 2-4: MEMPHIS distribution system

The User Database knows which document topics need to be delivered to which user in which format. It is consulted for each processed document, to perform the necessary tasks. The matching and prioritisation algorithms determine in which priority users will be informed, and which variant of the current documents needs to be prepared for sending. The output queue does the actual packaging, by providing a queue of documents which are ready for transmission, determining the output channel and wrapping each document accordingly. The output devices are components to support different output channels, like SMS, email, PDA's, and others. The number and kind of gateways depends on the applications; for some (like SMS) there is public domain software. The gateways do the real transmission.

In the current version, the MEMPHIS services provide output in the form of e-mails and SMS. Figure 6 and 7 demonstrate examples of the two type of messages.

MEMPHIS FinAlert Demonstration Service V.1

Co. / Org.: Northern Rock PLC, Lloyds TSB Group PLC

Title: Bad debt provisions give banks testing time

Date: Jan 28, 2003

Source: ft.com

Adam Applegarth, who heads the Newcastle-based bank, is expected to report pre-tax profits of between £316m and £325m, against £295.5m, helped by the booming housing market. "I think people will be looking at bad debt pro-visions, but if that hasn't deteriorated then I think the focus will switch quickly to sluggish revenue growth," said Alex Potter, at Lehman Brothers. Already Lloyds TSB and Barclays have cautioned that bad debt provisions in 2002 will rise because of their lending to large corporates, as well as to crisis-hit Argentina. Lloyds TSB cautioned about higher corporate bad debt provisions in 2002. The worst affected bank is Lloyds TSB, which has squeezed cost savings from past acquisitions. Bad debt provisions give banks testing time

Figure 4: Example MEMPHIS E-mail message

Figure 5: Example SMS message

4 Summary of Results from Internal Tests

Currently the core modules of the MEMPHIS system are applied for two demo information services: FinAlert and MediAlert. The FinAlert service delivers financial news on the topic "Merger and acquisition" world-wide. Original information and documents relevant for the topic are automatically collected from the online available free news provided by Financial Times (www.ft.com). The news are provided in two languages Italian and English. The news in Italian are generated through translation from originally English news. Than they are summarised and can be received on the users preferred output device as an SMS or as an e-mail.

MediAlert is a general push information service which offers information on various media products via individual distribution channels and output devices, and this in various languages. The documents are the announcements of publishing houses collected in the database of the publishing company Random House and Mount Media. English documents are retrieved from the online available information from Random House (www.randomhouse.com). German documents are retrieved from the online service of Mount Media (www.mountmedia.de). The testable prototype can deliver information about books and biographies related to the topic "Music" in two different formats: a short description (SMS) and a long description (e-mail) in two languages: in German or in English. Source information available in both languages is retrieved, checked for duplicate announcements and translated and summarised according to the chosen target languages of users.

Currently the two services are available online for 50 external test users. In the period from the 23th of November to the 20th of January the system was tested internally by employees of the developing companies. Thereby the following findings related to its functionality were collected:

- The acquisition systems relies in the current version on the keyword-search functionality offered by the source sites. Thus, the relevance of the retrieved documents differs from site to site depending on which search approach and technology is used. Sites as for example Random House provide better quality, as the information is classified manually. A special problem arose with detection of duplicates. Financial Times for example offers news categorised according to their actuality. The first time they are offered in the news section and than in the news

archive or as special reports. Thus, using the date of publishing as a duplicate detection criteria is not always successful. News related to the same event following up on the event during several days can have slight differences, which might get lost during summarisation in particular, when delivery is with SMS. In the next version of the services additional retrieval algorithms will be tested.

- The relevance of the retrieved documents with respect to the keywords has among other factors direct impact on the quality of translation and summarisation. Summarisation of free text documents as the news from Financial Times did not always result in meaningful SMS. This is partly due to the limited space of SMS messages. The ublisher's sites provide information in a more structured manner. As a result a structured SMS containing the basic information about media products as for example, title, author and similar could be extracted. It is expected, that in the next version of the services where PDAs will be added as possible output devices as well, summarisation will improve as longer messages then 160 characters will be possible.
- As the search is executed periodically, the delivery resulted in burst deliveries with several messages at the same time. In order to prevent this in future versions of the services, the possibility to control and define distinct delivery frequency will be implemented.

The two test applications of the MEMPHIS system demonstrated that multi-language applications have the following features:

- the more narrowed the information domain is the better translation and summarisation quality can be achieved;
- automatic generation of meaningful very short summaries for SMS delivery from large documents can not be guaranteed with sustainable quality, in particular where the source is unstructured free text;
- multi-lingual applications require a continuos adoption and development of the underlying vocabularies and ontologies for the specific domains.

This means that multi-language information services require extensive maintenance resources and will be suitable only for specific business scenarios. Given this the following potential business scenarios for the MEMPHIS system were identifies in two areas: as part of free or paid online information services or as part of enterprise information systems. In the context of online information services MEMPHIS can be applied as follows:

- Application of MEMPHIS as add-on to existing online services. Translation and summarisation as well as adoption to different kind of end devices can in particular enhance existing free online information services. MEMPHIS can be applied for charged premium information services to specific clearly defined topics and be offered to business customers willing to pay for such services. Thus, enabled by MEMPHIS the online information service can offer free services and paid premium services.
- MEMPHIS could also be used as an online premium information service itself by content syndicators.
- MEMPHIS services can also be part of advertising alert services as for example the test application MediAlert.

Commercial MEMPHIS online information services should support different revenue models: subscription based on flat fee for a certain period of time, payment per delivered message or combination of fixed subscription per defined period of time and variable payment per delivered message.

MEMPHIS can also be applied as a multi-language and multi-device enhancement of enterprise information system. In particular mobile workers as mobile

sales or repair and maintenance work forces of multi-national and thus multi-language companies can be empowered with targeted information services on mobile devices.

5 Outlook

We presented in this paper an overview of the multi-lingual in the MEMPHIS toolbox for retrieval and transformation of content for mobile end devices. The system is mostly based on state-of-the-art technologies, but combines them in innovative services which can empower very interesting business cases. Prototype systems for the services mentioned are already working. An end-to-end test phase of the services with "real" external test users is currently performed. At the time of the conference, we hope to be able to present evaluation results from this tests.

Acknowledgments

The work reported here is part of the MEMPHIS project funded by the European Commission under contract IST-2000-25045 in *the Information Society Technologies (IST)* program.

References

[1] Boston Consulting Group: Mobile Commerce: Winning the on-air consumer, 2000.

[2] Buyukkokten, O.; Garcia-Molina, H.; Paepcke, A.; Winograd, T.: "Power browser: Efficient Web browsing for PDAs. In: Proceedings of the Conference on Human Factors in Computing Systems, (CHI'00), 2000, pp 430-437.

[3] Buyukkokten, O.; Kaljuvee, o.; Carcia-Molina, H.; Paepcke, A.; Winograd, T.: "Efficient Web Browsing on Handheld Devices Using page and Form Summarization". In: ACM Transactions on Information Systems, Vol. 20, No. 1 January 2002, pp. 82-115.

[4] Eisele, A./Kasper, W. (under review): *Named Entity Recognition for Tuning Machine Translation Systems*. Under review.

[5] Kaasinen, E.; Aaltonen M.; Kolari, J.; Melakoski, S.; Laakko, T.: Two Approaches to Bring Internet Services to WAP Devices". In: Proceedings of the 9th Internationan World-Wide-Web Conference 2000, pp 231-246.

[6] Kasper, W./Steffen, J. (2002): *Multilingual flexible and robust summarization*. Proceedings Konvens-2002. Saarbrücken 2002, pp. 215-218

[7] M. Becker et al. (2002): *SproUT - Shallow Processing with Unification and Typed Feature Structures*. International conference on NLP. Mumbai 2002

[8] A. Yannopoulos, Y. Stavroulas, N. Papadakis, D. Halkos, T. Varvargou (2002): *A Method Which Enables the Assessment of Private Data by an Untrusted Third Party Using Arbitrary Algorithms But Prevents Disclosure of Their Content*. In Hamid R. Arabnia, Youngsong Mun (Eds.): Proceedings of the International Conference on Internet Computing, IC'2002, Las Vegas, Nevada, USA, June 24-27, 2002. CSREA Press, 2001, ISBN 1-892512-38-6 - Volume 2.

[9] Zobel, J.: Mobile Business und M-Commerce, München, Wien, Carl Hanser Verlag, 2001.