MITOS: An Integrated Web-based System for Information Management

Evangelia Kopanaki♦, Vangelis Karkaletsis♣, Constantine D. Spyropoulos♣,
Nikos Avradinis♦, Nikos Fakotakis♦♦, Theodore Kalamboukis*, Basilis Kladis+,
Yannis Lazarou♦, Themis Panayiotopoulos++, Diomidis Spinellis**

♣*Software and Knowledge Engineering Laboratory, Inst. of Informatics & Telecommunications,
NCSR "Demokritos", GR-15310 Ag. Paraskevi, Greece, e-mail: {vangelis, costass}@iit.demokritos.gr*

♦ *Kalofolias Group S.A., Kapa-TEL Information Network,
39 Halandriou Str., GR-151 25, Maroussi, Athens, Greece, {evik, avrad, lazar}@mail.kapatel.gr*

♦♦ *Wire Communications Laboratory, Dept of Electrical & Computer Engineering,
University of Patras, Patras, GR-26110, Greece, fakotaki@wcl.ee.upatras.gr*

* *Dept. of Informatics, Athens University of Economics & Business,
46 Kefallinias Str., GR-11251, Athens, Greece, tzk@aueb.gr*

+ *KNOWLEDGE S.A., NEO Patras-Athens 37, GR-26441, Patras, Greece, bkladis@knowledge.gr*

++ *Dept. of Informatics, University of Piraeus,
80 Karaoli & Dimitriou Str., GR-18534, Piraeus, Greece, themisp@unipi.gr*

** *SENA, Vyzantiou 2 Str., GR-14234, Nea Ionia, Athens, Greece, dds@sena.gr*

The wide availability and accessibility of information have made its management and deployment even more difficult. To this end, remarkable effort has been made for the development of information systems that handle the processing, analysis and management of information. However, the success of these systems does not only depend on the quality of information handling, but also on the appropriate presentation of information to the end-user. MITOS[1] system analyses financial news by employing techniques from the areas of Natural Language Processing, Information Filtering and Information Extraction. Moreover, by acknowledging the importance of the presentation of information, MITOS has also incorporated User Modelling techniques, which enable the provision of personalized content adapted to each user's profile.

1   Introduction

New technologies, such as high-speed networks and inexpensive massive storage along with the Internet expansion have led to a considerable increase of the amount and availability of on-line text. However, information is only valuable to the extent that it is accessible, easily

---

retrieved and structured [5]. The growing volume of data, the lack of structured information, and the information diversity has made its retrieval and management even more difficult [4]. Thus, there is a strong need for improved means of controlling the information explosion [13]. To this end, remarkable effort has been made towards the development of advanced techniques for organizing and accessing texts. These techniques include automatic document filtering [3] and extraction of selected information from on-line sources [2].

Aiming to address the problem of information overload, MITOS project developed an integrated system for information management able to filter and extract information from electronic news, derived from the Greek financial market. MITOS system combined techniques from the areas of *Information Filtering*, *Natural Language Processing* and *Information Extraction*. Moreover, by acknowledging the importance of the presentation of information, MITOS has also incorporated *User Modelling* techniques, which enable the provision of personalized content adapted to each user's profile. The co-operation of the aforementioned technologies provided an integrated system of information processing.

In order to better understand and satisfy user needs, a 1st prototype was created and evaluated. The feedback that we got enabled us to improve as well as add new functionalities to the system. Based on the users' remarks, a 2nd prototype of the integrated system has been developed and evaluated leading to the final system prototype.

In this paper we do not only present the technological features of MITOS project, but we also demonstrate the way the analysed information is presented to the end user. In Section 2 we present an overview of MITOS project, its objectives and initial user requirements, and we outline the system's architecture. In Section 3, we demonstrate the functionality provided to the end user, whereas in Section 4 we illustrate how the evaluation led us to better capture and satisfy users' needs. Finally, in Section 5 we present some concluding remarks.

2 System Overview

MITOS objective was to produce an integrated system for efficient information management exploiting the partners' expertise in the technologies involved, in the development of commercial applications as well as in the domain of finance (both from an academic and a commercial perspective). More specifically, MITOS aimed at providing:

➢ Information according to the users' interests and knowledge
➢ Automatic categorization of financial news in various thematic domains, enabling the user to select the thematic domains that are of interest to him/her, as well as to express specific queries in the financial news base using his/her own keywords.
➢ Automatic identification of financial news belonging to specific event types and extraction of their most important features. Thus, instead of having the user read the whole text of news, the system "reads" the news for him/her extracts the interesting features and provides them in a summarized format (a table).

In order to achieve the above objectives, MITOS project exploited and integrated the following technologies:

➢ *User modelling*: provides techniques for creating and maintaining users' models in order to provide and present personalized information to the user.
➢ *Information filtering*: provides techniques for categorizing texts in various thematic domains and retrieving texts matching a user's query. The MITOS information filtering sub-system developed and exploited statistical and machine learning based techniques taking into account the users' requirements concerning the thematic domains and the functionalities that should be provided.

- *Natural Language Processing:* provides the necessary infrastructure for the morphological, syntactic and semantic analysis of financial news.
- *Information extraction:* provides techniques for automatically extracting information from text and creating a structured representation of this information The MITOS information extraction sub-system contains tools that enable the recognition of named entities involved in the event, the identification of their roles (e.g. that the name entity A is the buyer company in a purchase event) as well as the extraction of other important information (e.g. period of time, amount of purchase). It takes into account the results of the information filtering sub-system in order to identify the news referring to a specific event (e.g. mergers, acquisitions).

For the development of MITOS system, we followed a *rapid prototype development* approach. This concerned both the sub-systems implementing the above-mentioned technologies as well as the integrated MITOS system.

During the 1st phase of the MITOS project [7]:
- financial experts specified the profile of the MITOS users, their information needs and the data required in order to cover these needs,
- the text bases of Kapa-TEL were studied in order to identify if they had to be updated to cover the users' needs,
- a comparative study of existing commercial financial applications (national and international) was performed focusing on the provided services and information types.
- financial experts specified thematic domains for information filtering and the event types for information extraction.

The results of the above tasks were used to devise the functional specifications of MITOS integrated system and sub-systems as well as to develop an experimental mockup that was presented to a user group [8]. The users' answers to a relevant questionnaire and their comments were taken into account for the development of the 1st version of MITOS sub-systems as well as for the development of the 1st prototype of the integrated system. The system's architecture and the integration of the technologies mentioned above are presented in the following section.

2.1 MITOS Architecture

MITOS system takes as input text of news belonging to various thematic domains, varying from finance to sports and politics (see Fig. 1). The Information filtering sub-system classifies news texts into predefined thematic domains in the field of economics and finance. The Information extraction sub-system takes as input the classified news, identifies those referring to specific events, extracts from them the interesting features and stores the results into the MITOS database. End-users interact with the system through a web interface. User interface functionality is enhanced by the user modelling sub-system that tracks user's actions and stores information about the user's interests and preferred search options.

MITOS is a distributed system. Its subsystems can reside in different servers and communicate via the TCP/IP protocol [8] (see Fig.1). In its current implementation, MITOS front-end, i.e. the user interface, resides in the main server under the Solaris Operating System. The user interface module (developed mainly using PHP, JavaScript and HTML) works under an Apache Web Server, installed on the same server. The MITOS database (PostgreSQL) containing the results of the information filtering and extraction sub-systems is also on the Solaris server. The MITOS back-end, i.e. information filtering, information extraction and user modelling sub-systems, have been installed on a Windows NT Server.

The User Modelling module works under Microsoft's Internet Information Server and communicates with the user interface module via HTTP GET actions.
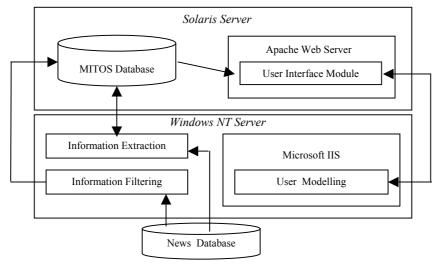

Figure 1. MITOS Architecture

The Information Filtering sub-system consists of a set of tools that acquire thematic domains' models from a news database (training corpus), and classify new texts into their corresponding domains [10]. More specifically, the Information Filtering Module consists of:
➢ A manual text classification tool necessary for the system's initial training.
➢ A statistical text processing tool that is used to extract keywords and key phrases from news texts, feeding the information filtering tool.
➢ An information filtering tool which automatically classifies the news text in the thematic areas supported by MITOS.
➢ Tools supporting the customisation of information filtering to new thematic domains.

The Information Extraction sub-system extracts important features from the classified news texts. This is performed using several processing stages [9, 10]:
➢ Named entity recognition, to identify the named entities involved in the texts.
➢ Named entity descriptions recognition, to identify phrases describing the named entities.
➢ Event recognition, to identify those texts describing one or more of the event types supported.
➢ Event roles recognition, to identify the roles of the named entities involved in an event as well as the other possible roles of this event.

The Information Extraction sub-system includes also tools supporting its customisation to new event types.

MITOS' User Modelling sub-system [6] keeps track of the user's selections each time he/she posts a search query, in order to determine the user's preferences and adjust some of the user interface's options according to these preferences. MITOS uses a long-term model approach, based on the assumption that user preferences are expected to appear after a period of system usage. Apart from creating a personal user profile, the user modelling sub-system also creates group profiles, the user communities [12]. A community corresponds to a group

of users who exhibit common behaviour in their interaction with the system. Model creation for personal profiles is performed on-line whereas the creation of user communities is performed as a batch process.

3. Functionalities Provided to the end users

The design of the user Interface was based on the users' requirements, specified in the first phase of the project [7]. Our objective was not only to exploit the information processing capabilities offered by MITOS technologies, but also to accomplish the desirable user satisfaction. We developed a web-based user interface, using JavaScript-enhanced HTML forms for query formation (visit MITOS web site at http://mitos.kapatel.gr). Considering that added value is not gained merely through large quantities of data on a site, but through easy access to the required information, we tried to design and develop a usable and friendly user interface that would facilitate the retrieval and management of analysed data.

Aiming to provide an intelligent data repository of analysed financial news, MITOS offers an efficient way of searching and retrieving the appropriate information. It allows users to exploit the categorization of news, as well as their presentation in a summarized form. Three different search modes are provided: simple search, advanced search and search based on the selection of companies' names.

The *"Simple Search"* mode enables users to look for news related to the categories of their choice (see Fig. 2). This way of searching is based on the classification of financial news, conducted by the use of Information Filtering techniques. Fourteen categories have been identified as the most representative thematic domains. Users can select the categories of their interests (e.g. mergers and acquisitions) and retrieve all financial news belonging to them. In order to limit the amount of information retrieved, users can specify additional terms of search such as a date or a period of time that they prefer. Moreover, users can further increase the accuracy of their search by forming complex boolean queries of keywords, using AND/OR/NOT/NEAR operators. This additional feature can be provided, since the analysis and processing of information enabled the determination of keywords characterizing the financial news. The simple search results are presented in a list that includes the date and title of financial news in a chronological order. The latter helps users to quickly identify and select to read the full text of news that interest them the most.

The *"Topic Search by Company Name"* mode enables users to select the thematic domains of news they prefer as well as one or more company names from a multi-selection list. In order to further limit their request users can also specify a date or period of time (see Fig. 3).

The "Advanced Search" mode enables users to search news based on specific event types, such as Acquisitions, Mergers, etc. (see Fig. 4). For each event type, they can define specific search terms and retrieve financial news in a summarized format. This operation exploits the results of Information Extraction, which analyses news, belonging to the aforementioned events, and extracts from their content the most significant elements of information. The analysis and identification of the most important data enabled not only the presentation of news in a summarised format, but also the use of multiple criteria for the provision of a more efficient way of search
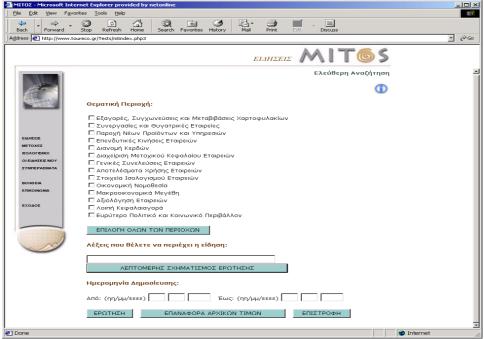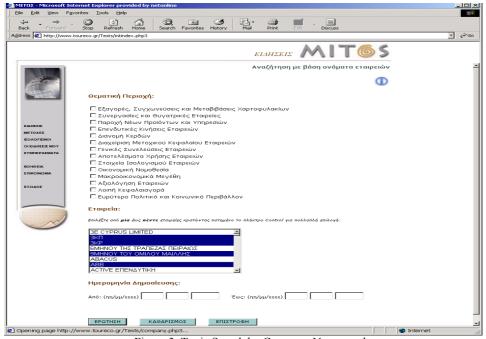
Figure 2. Simple Search mode

Figure 3. Topic Search by Company Name mode

Figure 4. Advanced Search mode

In order to use the "Advanced Search", users need first to select a specific category (e.g. Acquisition, Merger). According to the event type selected, the search form is dynamically modified in order to present a set of search terms fit to the selected event. These terms of search are distinct to each event type. For example, when users search for news involving a company's acquisition (see Fig. 5), they can specify the names of the companies involved, the percentage of acquisition, and the period of time that interests them, whereas when they are looking for a merger, they can only specify the involved companies and the period of merger. The news retrieved, are presented in a summarized format, which incorporates the most significant information (e.g. Date, Company name, amount of money exchanged, percentage of acquisition and time of acquisition). The users can select to view only the parts of information that interests them. They can also select the ordering they prefer (e.g. by date or by percentage of acquisition). Finally, users can access the full text of news by simply selecting their date from the list.

All these functionalities were enhanced adopting user-modelling techniques to provide personalized content to the users. In particular, we focus our work on the concept of a user community, which seems to apply best to a public Web site. The system monitors and understands users' preferences regarding specific companies and thematic domains. By maintaining the users' needs the system can provide personalized user interfaces. This feature facilitates the use of both "simple" and "advanced" search, by pre-selecting the thematic domains that interest the users.

Another feature offered by MITOS system is the "Personal Newsletter", which presents the most recent news according to the user's profile. The user-modelling module provides information regarding the user's most preferred thematic domains and the user's most preferred companies. It then displays the full text of all recent news related to the preferred domains and companies.
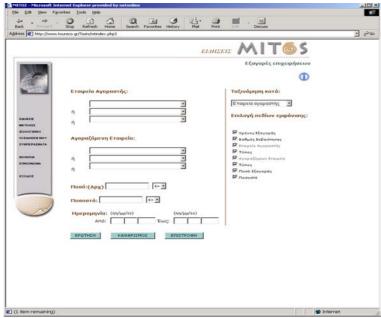
Fig. 5 Advanced Search mode (company acquisition event)

4. Evaluation

A *Rapid Prototype Development* approach was applied in MITOS. The evaluation of prototypes was conducted in two levels: testing of the back end sub-systems, and testing of the integrated system. The efficiency and acceptance of the integrated system was examined through the evaluation of the user interface and its functionality.

The first level of system evaluation is described briefly, since it is beyond the scope of this paper. For the testing and further improvement of the Information Filtering, Natural Language Processing, Information Extraction and User Modelling subsystems, multiple prototypes were created and evaluated during the course of MITOS project [9, 10, 11, 6]. The development of prototypes facilitated early testing and provided a common reference point for all members of the design team. It improved the quality and completeness of the subsystem's functional specification and led to the identification and solving of problems at an early stage of the system's development.

Having realized that user requirements frequently do not become apparent until a system is in use, we tried to provide, from an early stage, a common base of discussion between developers and users. We, thus, developed and tested multiple prototypes of MITOS front-end. The evaluation of the user interface and its provided functionality was conducted in the following three stages: creation of a mockup, development and evaluation of a first prototype, development and evaluation of the second prototype

The mockup consisted of HTML pages, which presented the main functions of the system. In that early stage MITOS back end had not been developed. Our goal was to present a preliminary design of the site's functionality. Therefore the data presented in the site were just examples especially created for the needs of the mockup. The usefulness of the mockup

evaluation was rather limited. Instead of looking at the presented functionalities, most users claimed that they could not give valuable feedback since the data provided was limited and the supported functions incomplete. Therefore we had to wait the development of the first prototype, in order to examine the level of system' s acceptance as well as to better understand the users' needs.

The first prototype enabled us to show something tangible to the users, before committing them to the final system. Prototypes usually represent an incomplete version of the final system consisting of inputs, intermediary stages and outputs [1]. In our case, we were able to provide some level of functionality, since preliminary versions of back-end systems were released. We, thus, incorporated a large amount of financial news, which were processed off-line by the Information Filtering and Information Extraction sub-systems. The analysed information was stored in the system's database and thus became available to the end user. The first prototype supported only two types of search: Simple and Advanced. It also incorporated user-modelling techniques for the provision of personalized content. We, thus, presented to users a prototype supporting the main MITOS functions. The evaluation was conducted by persons interested in financial business as well as by persons experienced in the development of information processing systems. A brief description of the system's functionality along with some suggested usage scenarios were distributed to the evaluators. Since our aim was to assess whether MITOS system satisfied the users' needs the data collected was mainly qualitative.

The users commented that the functionalities provided by MITOS system were useful. By being able to test the system using real data they could better appreciate its utility. However, they requested major improvements to the user interface, which seemed to be poor and unattractive. The limited attention that was given to the design of the user interface prohibited users of appropriately assessing the system's functionalities. They also mentioned, the lack of a detailed online help. Besides the improvements in the MITOS web site appearance, the users also suggested to incorporate some additional functionality. Although they were satisfied from the provision of two ways of search, they recommended adding the feature of searching by using company names. They also proposed the provision of an extra Web page, which would include the most recent news of interest, in the form of a personal newsletter.

Following users' suggestions, the third way of search, called "Topic Search by Company Name" was incorporated. We also offered the "Personal Newsletter" service to present the most recent financial news related to thematic domains and companies of users' interests. Besides adding these features, we have also made major improvements to the overall appearance of the user interface. In order to provide a friendly and attractive environment we designed some Web pages asking for feedback from the users. The agreed web pages were incorporated to the final version of the system. We finally enhanced our user interface by including a more detailed online help.

In the second prototype emphasis was not only given to the evaluation of its operations but also to the quality and accuracy of the processed information. A beta version of the Information Filtering and Information Extraction sub-systems was installed in the MITOS server. We then started to feed our database with financial news, which were analysed and processed online by the back end's systems. During the evaluation of the 2nd prototype, users reported only minor weaknesses and mentioned limited problems related to the systems functionality. Most improvements suggested were implemented in the final version of the system.

Besides these limited remarks the overall system was found to be useful. According to users' opinion, the second version of the user interface is easy, attractive and friendly to use. This enabled them to pay more attention to the evaluation of the content and functionality of the site. Most users were satisfied by the results of the analysis and processing of information and argued that MITOS system facilitates the retrieval of financial news. They also said that it enables them to easily and quickly identify valuable information.

5. Conclusions

In this paper, an overview of the MITOS system was presented. MITOS provides an integrated environment for information management. Emphasis was given to the development and integration of the Information Filtering, Natural Language Processing and Information Extraction tools. The satisfaction of the end user was also taken into account incorporating User Modelling techniques that enable the provision of personalized content to the users. This R&D effort was realized following a rapid prototype development approach. This approach facilitated us to identify problems at an early stage of the system development as well as at the integration phase, satisfying at the same time the end user and improving the quality and completeness of the provided functionalities.

References

1. Avison D.E. and Fitzgerald G., "Information systems development: methodologies, techniques, and tools", 2nd ed. New York: McGraw-Hill, 1995.
2. Cowie J. and Wilks Y., "Information Extraction", Handbook of Natural Language Processing, ed. R.Dale, H.Moisl, H.Somers, Marcel Dekker, Inc. 2000, pp. 241-260.
3. Hanani U., Shapira B. and Shoval P., "Information Filtering: Overview of Issues, Research and Systems", Journal of User Modeling and User-Adapted Interaction, vol. 11, no. 3, pp. 203-259, 2001.
4. Kalakota R. and Whinston A. B., "Frontiers of electronic commerce", Addison-Wesley Pub. Co, 1996.
5. Koniger and Janowitz, 1995 "Drowning in information, but thirsty for knowledge", International Journal of Information Management, Vol. 15, No. 1, pp 5-16, 1995.
6. MITOS project (NEO EKBAN - 102), Deliverable 7.1 "User Modelling – Technical report", March 2001.
7. MITOS project (NEO EKBAN - 102), Deliverable 2 "Users' Requirements", January 2000.
8. MITOS project (NEO EKBAN - 102), Deliverable 4 "Functional Specifications and Architecture Design", March 2000.
9. MITOS project (NEO EKBAN - 102), Deliverable 5.1 "Language Resources and Tools – Technical report", March 2001.
10. MITOS project (NEO EKBAN - 102), Deliverable 7.1 "Information Filtering – Technical report", March 2001.
11. MITOS project (NEO EKBAN - 102), Deliverable 8.1 "Information Extraction – Technical report", March 2001.
12. Paliouras G., Papatheodorou C., Karkaletsis V. and Spyropoulos C.D., "Clustering the Users of Large Web Sites into Communities". Proceedings Intern. Conf. on Machine Learning (ICML), Stanford, California, 719-726.
13. Rowley J., "Towards a Framework for Information Management", International Journal of Information Management, Vol. 18, No. 5, pp 359-369, 1998.