Report on The fourth International Conference on Flexible Query Answering systems (FQAS 2000)

Patrick Bosc*, Amihai Motro+, and Gabriella Pasi**

*** IRISA ENSSAT, Lannion, France.**
Patrick.Bosc@enssat.fr
+ George Mason University, USA.
ami@gmu.edu
**** ITIM – CNR, Milano, Italy.**
gabriella.pasi@itim.mi.cnr.it

1 Introduction

The fourth International Conference on Flexible Query Answering Systems (FQAS'2000) was held at the Academy of Sciences in Warsaw, Poland on October, 25-27, 2000. This series of conferences was launched in 1994 by Troels Andreasen, Henning Christiansen and Henrik Larsen from Roskilde University in Denmark, who have been the main driving force behind this series ever since. The previous FQAS events were held in Denmark in 1994, 1996, and 1998. The next conference in this series will return to Denmark in 2002.

This conference provides an opportunity for researchers, developers and practitioners to explore new ideas and approaches in a multidisciplinary forum to the aim of providing effective solutions to the problem of easy, flexible and intuitive access to electronically stored information. The focus of this conference is to add flexibility to the systems for the storage and access to information; as there exist several classes of such systems, the conference draws on several research areas, including information retrieval, database management, knowledge representation, multimedia systems and human-computer interaction.

As a paradigm for flexible query answering we may think about an expert human intermediary who is able to analyze users information needs and to evaluate the relevant information items from the available information sources. The knowledge on the information sources and the capability to interpret the user requests enable the expert to perform a good estimate of the items possibly satisfying the users needs, though the query, per se, may be imprecise, incomplete, etc. Thus, one of the key issues for defining flexible query answering system is the tolerance to imprecision and uncertainty in the formulation of user queries as well as in the representation of information.

The conference brought together more than 70 researchers and developers working in the area of information technology. The conference program included five invited talks, a keynote speech and 12 technical sessions. There was a total of 55 reviewed papers included in the technical sessions.

In the proceedings, the contributions are gathered according to 6 main topics, each one corresponding to one or more technical sessions. These principal themes are the following: Flexibility in Database Management and Querying, Information Retrieval, Knowledge Representation and Reasoning, Natural Language and Multimedia Processing, Acquisition of Data and Knowledge, Cyberspace and distributed data.

The rest of this report provides a brief summary of the three main thematic areas.

2 Flexibility in Information Retrieval Systems

With the increasing production and exchange of multimedia information through the Internet, the need for effective information retrieval systems is a crucial issue nowadays. Information retrieval aims at modeling, designing and implementing systems able to provide efficient and effective content-based access to a large amount of information. Information can be textual, visual and auditory, although most of the existing IRSs are able to manage only textual information. The main aim of an information retrieval system is to identify the information relevant to the needs expressed by a user thanks to a formal query (an expression of a formal query language). To this aim, both documents and queries are represented in a formal way and a so-called matching mechanism compares the two representations for estimating the relevance of documents with respect to the considered query. IR is a complex decision making activity, pervaded with vagueness and uncertainty, mainly due to the fact that an interpretation of both the users information needs and the document information content is needed. Most of the existing information retrieval systems and search engines are based on simple models to the aim of privileging efficiency at the expenses of effectiveness. A promising direction to improve these systems is to make them flexible, i.e. capable to be tolerant to imprecision, partial truth, uncertainty and approximation which characterize various stages of the IR process and to make them able to learn the user concept of relevance.

Formal theories and techniques such as probability theory, fuzzy set theory, evolutionary computing and neural networks are successfully employed to define flexible IRSs. At the FQAS conference several papers have proposed to contribute to the definition of flexible IRSs. Distinct aspects of information retrieval have been taken into account among which: indexing, query language definition, evaluation, natural language processing, etc.

Several papers address the problem of defining flexible query languages for IRSs. The problem of spoken query processing has been addressed by two papers; in the paper by Crestani, the results of an experimental study are reported, which show the effects of word recognition errors in spoken queries on the effectiveness of a spoken query processing system. Ballim and Pallotta describe a grammar-based method for a semantic analysis of dialogues; this method allows automated telephone-based phone-book querying. Quaresma and Pimenta Rodrigues describe their work on natural language interfaces for information retrieval systems. They propose a collaborative dialogue framework that is able to collaborate with users in order to refine their queries.

The problem of video retrieval is faced in the paper by M. Detyniecki, who proposes a tool for leading users to express questions aimed at the retrieval of information inside a video. The paper by Andreasen, Nilsson and Thomsen introduces the main objectives of the “Ontology-based Querying” project; the authors present an approach to text retrieval by means of domain ontology-based descriptions of both source texts and queries. Lalmas, Rolleke, Turra and Fuhr focus on Hypermedia Retrieval and propose a graphical user interface for querying hypermedia documents.

There has been considerable interest recently in the representation and retrieval of document written in XML or in languages for defining semi-structured documents. Because the structure of a document is generally related to the semantics of its information content, the exploitation of this structure in the indexing and retrieval phase is critical for flexible and effective access to information. Several papers at the FQAS meeting address issues related to retrieval from archives containing semi-structured documents. These include papers by Bordogna and Pasi, Hacid and Toumani, and Bellettini, Damiani and Fugini.

Some papers address the important problem of textual document indexing. Baranyi and Koczy propose an approach to text indexing that is based on neural networks. Kraft & Chen address the problem of document clustering in IRSs. Fuzzy

clustering algorithms are applied to classify documents, and fuzzy logic rules are employed for finding associations between terms, to the aim of deriving query modifications for a more effective retrieval of information.

Aguiar and Beigbeder propose an indexing procedure of WWW sites based on the definition of a 2-level index. This procedure takes into account the context of the pages local to WWW sites, and constitutes the basis of a search engine also proposed by the authors. Enguehard addresses the problem of term recognition in documents, and proposes an operator of flexible-equality of terms, which determines whether two terms can be considered as variations one from another.

Martin-Bautista, Sánchez, Vila and Larsen propose some fuzzy measures for evaluating the retrieval effectiveness of IR systems.
The paper by Cardenosa, Iraola and Tovar concerns another interesting problem related to IR, namely the information extraction problem; in their paper the authors address the problem of the extraction of names from newspapers articles.

3 Flexibility in Database Management Systems

Contrary to information retrieval systems, database management systems are founded on the use of a data model (relational or object-oriented) in order to structure the data of the universe concerned by a given database. In this context, no semantic extraction has to take place and the data are exploited as they are stored in the system. However, this difference does not preclude some room for flexibility in database systems and this term can have three major understandings. The first one is related to the queries that can be posed by users. Usually, queries are intended for searching items which satisfy a Boolean condition and a system will be said flexible if it allows to express preferences (and/or importance of criteria) in order to rank the retrieved items (which are more or less acceptable) according to their adequacy. Different techniques such as distances or fuzzy sets can be selected as the basis for defining preference mechanisms. Another aspect of flexibility in database systems concerns the nature of the data managed by the system which can be "imperfect" or "imprecisely known" and this point can be seen as a kind of refinement of null values. Here again, different frameworks and tools can be used to represent this kind of information among which or-sets, probability theory or possibility theory. Of course, flexibility can be understood in many other ways, such as the capability for a system to answer a query in a cooperative fashion.

The papers presented at FQAS'2000 that deal with these various aspects are now briefly overviewed.

Five papers are centered on the handling of imperfect data in relational or object-oriented frameworks and some of them consider flexible queries against such databases. The paper by Blanco, Marin, Pons and Vila addresses the issue of extending the data definition language of a relational DBMS in order to allow for the description of imperfect data represented as possibility distributions. The paper by Bosc, Li? tard and Pivert reviews some semantics of the join operation in presence of imperfect data. The work of Morris and Jankowski is related to geographic information systems where they suggest to take into account imperfect data and to provide users with a fuzzy query language. The paper by Koyuncu, Yazici and George proposes a system in which both the database and knowledge components can deal with imperfect data which are manipulated thanks to fuzzy predicates. Buche and Hæmmerlé report on a more applied work, namely an application on microbial risk in food products where imprecise data are stored and queried using fuzzy conditions.

In the two following papers, the authors use flexible queries for specific purposes. The work of Zbroja and Ligeza is situated in the area of case-based reasoning. They propose an extension of a relational system in order to support a flexible matching mechanism required by case-based reasoning. Kacprzyk and Zadrosny address the issue of discovering linguistic association rules which represent an extended version of regular association rules (in the sense of Agrawal and Srikant) where linguistic terms replace elementary values. The paper shows how such rules can be

mined thanks to a flexible (fuzzy) query language (FQUERY) formerly designed by the authors.

The papers by Demolombe on the one hand and Bidault, Froidevaux and Safar on the other hand are concerned with cooperative answering, i.e., systems in which the user is provided with an answer which is not necessarily the one induced by the query itself. The second one is situated in the context of a system made of the integration of several independent databases.

4. Information Integration and Inconsistency; Knowledge Discovery and Data Mining.

Issues of information integration and inconsistency were the subjects of several papers. Using the framework of logic, Arenas, Bertossi and Chomicki show how to query a relational database so that only consistent data is retrieved. Motro, Anokhin and Berlin address the issue of inconsistency within systems that integrate information from multiple sources and suggest several approaches based on voting and fusion, while taking into account the quality of the information. In addition, they argue that the manual mapping of new sources into an integrating system is a costly and daunting task, and propose a method for the automatic discovery and mapping of new content. A related issue of query planning in systems that integrate data from multiple sources is addressed by Lesser and Naumann, who suggest speeding up this expensive task by considering information quality. The integration of multiple, independent information sources is also the subject of a paper by Bidault, Froisevaux and Safar, who develop methods of repairing global queries that could not be translated properly over the available set of information providers. In the related field of data warehousing, Albrecht, Hummer, Lehner and Schlesinger show how semantic information can be used to improve the process of query rewriting, in which OLAP queries are rewritten to use the available materialized views (usually, summary tables) .

There were several papers in the areas with knowledge discovery and data mining and we mention here briefly five contributions. Alcamo,Domenichini and Turini present an XML-based environment for knowledge discovery, in which complex knowledge extraction problems can be expressed. Boulicaut, Bykowskei and Jeudy focus on the tractable discovery of association rules with negation. Greco, Masciari and Pontieri describe a system that integrates different data mining techniques, including deductive tools, data mining tools and visualization tools, to guide users in the data analysis process. Kacprzyk and Zadrozny discuss the mining of linguistic association rules using a flexible (fuzzy) querying interface. The inappropriateness of standard query languages for situations in which the information needs of users are vague has been known for a long time. Kaufman and Michalski observe that a similar situation occurs in the process of knowledge discovery, when users find it hard to define the target knowledge precisely. To address this issue, they propose the mechanism of »knowledge scouts» which are software agents that automatically synthesize knowledge of interest to a given user by applying inductive operators to the data sources.

5 Conclusions

In this report we have summarized the three main thematic areas of the fourth International Conference on Flexible Query Answering Systems (FQAS 2000), i.e. Information Retrieval, Database Management and Knowledge Discovery and Data Mining. The high quality of the scientific contributions and the considerable attendance in the event itself are evidence of the extent of interest in flexible systems for the management of information. Because of the ever-growing use of electronic information by non-professionals, this interest has been increasing steadily since the first FQAS (1994) and is expected to keep increasing in the foreseeable future.

The proceedings of the conference were published as a volume in the series “Advances in Soft Computing” of Physica Verlag, with the title “Flexible Query Answering Systems: recent advances” (editors: H.L. Larsen, J. Kacprzyk, S. Zadrozny, T. Andreasen and H. Christiansen).