Report on FQAS 2002: Fifth International Conference on Flexible Query Answering Systems

Amihai Motro
George Mason University
Fairfax, VA, USA

Troels Andreasen
Roskilde University
Roskilde, Denmark

The fifth conference on Flexible Query Answering Systems took place from October 27 to October 29, 2002 in Copenhagen, Denmark.

The overall theme of the FQAS conferences is *innovative query systems* that are aimed at providing easy, flexible and intuitive access to information. Such systems are intended to facilitate retrieval from information repositories such as databases, libraries, and the World Wide Web. These repositories are typically equipped with standard query systems, which are often inadequate, and the focus of FQAS is the development of query systems that are more expressive, informative, cooperative and productive.

The subject is inherently multidisciplinary, drawing on many research areas, including database management, information retrieval, artificial intelligence, knowledge discovery and data mining, knowledge representation and ontologies, soft computing, classical and non-classical logics, computational linguistics and natural language processing, multimedia information systems, and human-computer interaction.

The first FQAS conference was in 1994, and the conference has been a biennial event ever since. The sixth conference will take place in Lyon, France on June 24-26, 2004 (http://fqas2004.org).

The 2002 meeting featured three keynote lectures. The first keynote lecture, "Crossing the Structure Chasm," was presented by Alon Halevy. It observed that online information is of one of two distinct kinds: structured information (e.g., databases) and unstructured information (e.g., text). This distinction has lead to the development of two disparate "worlds" (paradigms for authoring, sharing, searching, and so on). The speaker presented Revere, a system under development in the University of Washington, that aims to cross this chasm.

Researchers in the area of possibilistic (fuzzy) databases were among the earliest to provide solutions in support of flexible querying, and the second lecture, "On Different Families of Queries Addressed to Fuzzy Databases," by Patrick Bosc, was in this area. In this talk, the relatively broad term *flexible query answering* was assumed to have a particular interpretation: the ability to store, manipulate and retrieve imprecise information, and the talk surveyed various possible approaches to imprecision within the possibilistic framework.

In the third lecture, "Description Logics for Information Modeling and Access," Enrico Franconi, discussed yet another chasm: the one between the area of knowledge representation (specifically, ontology design), where powerful deduction capabilities have been applied to relatively small data sets, and that of database or document retrieval systems, where the emphasis has been on efficient handling of large amount of information, but with only limited reasoning capabilities. The speaker argued that description logics can be used to provide an integrated approach to knowledge representation and databases.

The main part of the program consisted of a selection of 29 papers. The papers addressed a wide range of topics, but several themes seem to have emerged. In this report we briefly discuss three of these themes.

There were several papers on *information integration*. Research in this area is concerned with the integration of distributed information sources that are possibly heterogeneous and mutually inconsistent. Badea and Tilivea presented a paper that models query planning (the decomposition of a global query to queries against the underlying sources) as a constraint handling problem. In another paper, Bertossi, Chomicki, Cortés and Gutiérrez addressed the issue of mutual inconsistencies in independent sources (with respect to global integrity constraints), and intro-

duced a methodology for generating query plans that retrieve consistent answers. A related problem of distributed inconsistency was addressed by Muñoz-Escoi, Irũn-Briz, Galdámez, Bernabéu-Aubán, Bataler, Bañuls and Decker. The issue here is that replicated information in a distributed information system can be temporarily inconsistent, causing a tradeoff between correctness and availability of answers in a networked environment. Querying multiple databases was also the subject of a paper by Parimala and Kumar. The authors propose a query language which is structure-independent (has no reference to the structure of the databases), thus isolating users from the need to include specific "path expressions."

Various techniques in the field of flexible query systems may be considered "classical", and there were several papers that suggested new approaches to the problems addressed by these well-established techniques. *Query expansion* is a method for transforming user queries to broader queries, thus increasing recall. Akrivas, Wallace, Stamou and Kollias proposed using fuzzy clustering of index terms to expand queries in an information retrieval system. *Interactive query construction* is a technique that assists naive users in constructing proper queries. Kapetanios and Groenewoud described a query construction method that emphasizes the suggestion of meaningful query terms to users. Interactive query construction was also the subject of a paper by Trigoni. In the method suggested, users begin with an expression of high-level (general) interest in a subject, and then "drill-down" to more specific queries based on answers received. The environment is that of semistructured databases and the process culminates in specific queries in a language such as XQuery.

The difficult problem of determining whether two elements (data values, terms, documents, etc.) are *similar* (as opposed to *identical*) is indigenous to flexible query systems, and several additional papers addressed it in one form or another. Nearest-neighbor (vague) queries are queries in which a specific multi-feature data object is presented to a repository, and the "best matches" are sought; that is, data objects that are closest to the given object, as calculated by a preferred distance metric. Dang, Küng and Wagner introduce a new algorithm said to be suitable for application domains where error is more tolerable. Bulskov, Knappe and Andreasen discussed methods for measuring similarity between concepts using ontologies. Classical information retrieval is concerned with ranking documents according to their similarity to a given specification. Ciaccia and Penzo ad-

dressed the implications of having XML structures in the searched documents.

In addition to the aforementioned papers, papers were presented on a wide variety of other subjects, including the fusion of ranked sets (Lalmas); an analysis of queries that indicate both negative and positive preferences (Dubois and Prade); a visual tool for querying time series data (Keogh, Hochheiser and Shneiderman); a model for querying the spatio-temporal properties of video streams (Koprulu, Cicekli and Yazici); and a knowledge-based query system for biological databases (Bresciani and Fontana).

The full list of accepted papers, acknowledgment of the individuals involved in organizing this conference, and other pertinent information may be found at http://www.fqas2002.org. The complete proceedings of the Fifth International Conference of Flexible Query Answering Systems are available from Spring-Verlag (Volume 2522 in the Series Lecture Notes in Artificial Intelligence).