

FUZZY LOGIC APPLIED TO INFORMATION RETRIEVAL SYSTEMS

Dulce Magalhães de Sá¹

Abstract — *Information retrieval systems can be used in a variety of contexts. For example, these systems can be used to support business or learning environments and digital libraries. Information retrieval systems are a type of resources that allow to obtaining information for decision making. Some reasons for information retrieval problems are data structures issues, efficiency of interfaces, information organization, database management systems absence, quality of information and information systems not suitable for specific process use. As a form to reduce the number of problems in information retrieval, fuzzy applications can be used to detect partial needs of user queries and access information element. It also can be used to organize the system answers by ranking of importance to facilitate the user decision, because in fuzzy logic an information element can reside in more than one set of different degrees of similarity.*

Index Terms — *Digital Libraries, Fuzzy Logic, Information Organization, Information Retrieval Systems, Multimedia Applications, Retrieval Techniques.*

INTRODUCTION

One important topic for design and develop information retrieval systems is the data structure and organization of information contents. Data structure is data grouped by an organization method of simple data collections [1]. Between the possible types of data structures are records, sets or files.

The evolution of the Internet, particularly its World Wide Web component, has created new opportunities and ways of application and analysis of information retrieval systems. One of the components of information retrieval systems is information itself that, in certain ways, has become a social and economical product.

One of the ways of organising information is by creating databases, developing applications and implementing database management systems [2]. Interactive databases in information retrieval systems are a critical component of the success of the system and, eventually, of the organisation that owns it, because they constitute a decision support tool.

INFORMATION RETRIEVAL ISSUES

Information retrieval is the ability of obtaining answers on information needs expressed to system in a language, usually a query language. Information retrieval systems are developed to provide efficient ways of search. Reference [3] defines information retrieval system as a system used to

store items of information that need to be processed, searched, retrieved and disseminated to various user population.

The basic components of retrieval information systems are search formulation, search software, information storage environment and queries process [4]. This involves hardware for information storage and persons (searchers) that executes queries.

- *Searcher* - person who has information or information contents needs and then begins the search process. In a particular case, the searcher can be a mixed entity between any person who elaborates part of the initial search process and a software agent who limits the scope of the search [2]. The human search is mostly oriented by subjective processes. The software agent searcher transmits to system the subjective aspects of the human user through rules previous fixed. Software agents can interfere in other phases of search, such as structures of search formulation [5].
- *Search formulation* - a process that requires some decisions that concern to topics of search, information fonts or contents, design of search formulation and which resources should be used in search process.
- *Resources* - software to search in local mode, private networks or Internet [4].
- *Information storage* - an important aspect of retrieval information systems because it allows search across the data structures.
- *Retrieval items* - aspects to performing the way that the system answers to queries depends of its design and conception.

Retrieval techniques are methods or processes used by systems for extract information topics in a particular way. A classification of retrieval techniques [6] distinguish between exact and partial match techniques. Exact match, for example Boolean retrieval and string matching techniques, requires precise information contents concern to topics of search.

Partial match techniques are used to compare queries with information elements represented as sets of features or index terms. Its include techniques based on formal models of information retrieval as probabilistic model, based on the probabilities of relevance to the query or fuzzy set model.

BOOLEAN INFORMATION RETRIEVAL

Information retrieval systems normally accept Boolean expressions, based on logic operators, as queries. Logic

¹ Dulce Magalhães de Sá, ISEGI – Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal, dulce@isegi.unl.pt

operators provide to their subject elements (Boolean data type), operations of conjunction, disjunction and negation. Negation operations are processed as exclusion of particular elements in a set of various elements. Logic operators are AND, OR and NOT [7].

Logic operators are used to clarify search formulation in retrieval systems. Let A, B and C words (strings) of any search formulation [A AND B] provided as answer all records with A and also B.

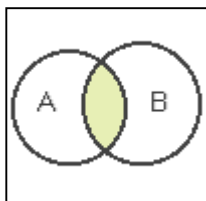


FIGURE 1
BOOLEAN OPERATION: [A AND B]

[A OR B] provided as answer all records with A, all records with B, and also all records with A and B.

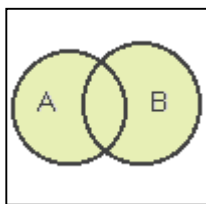


FIGURE 2
BOOLEAN OPERATION: A OR B

And [A AND B AND (NOT C)] provided as answer all records with A and B, but not those with of A and of B but not C.

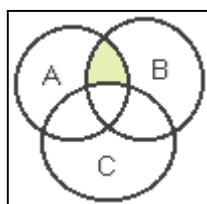


FIGURE 3
BOOLEAN OPERATION: A AND B AND (NOT C)

To obtain efficient results some design aspects must be considered for developing information retrieval systems. They must have a user-friendly search environment, including interfaces and data structures with Boolean type which supports logic operators [8].

However, Boolean information retrieval no ranking answer elements. The way that information contents are treated by retrieval systems and the way that it is presented as an answer to a query or request, are important factors to decision make process.

Difficulty to obtain answers to specific queries and access difficulties to found information contents or valid references to information needed are typical problems on information systems [9] and retrieval systems. A retrieval based on fuzzy set model provide searches with a decision rules to filtered information on databases and can be ranking answer elements to facilitate the decision make process.

FUZZY INFORMATION RETRIEVAL

Fuzzy logic is basically a multi-valued logic that allows intermediate values to be defined between usual valuations like 1/0, yes/no or true/false. It has been used initially in system theory to describe and implement uncertain notions and general concepts [10].

In fuzzy logic an information element can reside in more than one set of different degrees of similarity. Fuzzy relations represent a degree of presence or absence of association, interaction or interconnectedness between the information elements of two or more fuzzy sets.

One of the components of a fuzzy logic system is rules. These rules will be expressed as logic restrictions, in the forms of IF-THEN statements [11]. They are usually of a form similar to the following: *if x is low and y is high then z=medium.*

The area of information retrieval has also benefited from fuzzy logic methodology. Some fuzzy operations like union, intersection or complement can filter information elements with values between absolute true (1) and absolute false (0).

This provides a tool for natural language interfaces in retrieval systems and solution for some information retrieval problems. The software agent searcher that transmits to system the subjective aspects of the human user through rules previous fixed can be based in a fuzzy logic structure [2].

There already exist query systems that provide an ordering among the information items that more or less satisfy the request. The systems may allow for the presence or imprecise, uncertain or vague information in the system database.

One considers a system where the domain of each expression X (word or set of words) is represented as a function of a theme Y. If the expression X is associated to theme Y, X is relevant else X is irrelevant as retrieval item.

The grade of relevance is express for the weight of each X in a Y by numbers: 3 for must relevant, 2 to relevant and 1 to irrelevant. X is must relevant to theme Y if it is associated to weight 3, relevant if the weight is 2 and irrelevant for grade 1.

Weight 3 can be associated to a probability of X is include in Y, for example, between 1 and 0.75; for weight 2, if between 0.75 and 0.5; and for weight 1, from 0.5 to zero. This reduces retrieval items associated to theme Y because only consider "relevant" weights 3 and 2, supposing a rule which determine relevance as a probability of 0.5 as minimum.

CONCLUSION

Where information elements are stored in systems database, they can be classified by one of three weights, but this depends of the subjectivity of person who classified that. This process can be applied in simple and specific systems.

For large and complex systems is more reasonable that decision process depends of searcher. Thus, the retrieval system provides to searcher an optional theme Y for expression X gives by searcher.

System selects in information elements of database or in documents which X_i had weights 3 and 2 on Y_i and retrieved them. X_i and Y_i are any X and any Y respectively in each record on database or document (file) where X or Y exist.

System can be incremented, for example, by more variables like one second expression Z or more weights for relevance or other elements. This gives more complexity to system processing, but easiness to searcher person that use system and provide more efficient answers.

REFERENCES

- [1] Sá, D. M., Ferreira, M.C. and Santos M.J., *Tecnologias*, Porto Editora, Porto, PT, 1996.
- [2] Sá, D. M., "Fuzzy Web-based Information Retrieval Systems", *Proceedings of International Conference on systems Science*, World Scientific and Engineering Academy and Society, Oct 2002, pp. 1281-1284.
- [3] Salton, G. and McGill, M., *Introduction to Modern Information Retrieval*, McGraw-Hill, New York. USA, 1983.
- [4] Large, A., Tedd L.A., and Hartley, R.J., *Information Seeking in the Online Age: Principles and Practice*, Bowker-Saur, London. UK, 1999.
- [5] Pitkow, J., "In Search of Reliable Usage Data on the WWW", *Proceedings of the 6th International World Wide Web Conference*, 1997. Available on <http://www.scope.gmd.de/info/www6/technical/paper126/paper126.html>
- [6] Belkin, N. and Croft, W., "Retrieval Techniques", *Annual Review of Information Sciences and Techniques*, Vol. 22, 1987, pp. 109-145.
- [7] Hein, J. L., *Theory of Computing*, Jones and Bartlett Publishers, Massachusetts, USA, 1996.
- [8] Magalhães, D., *Interfaces com o Utilizador em Sistemas de Informação Web*, Working Paper n°. 66, ISEGI, Universidade Nova de Lisboa, Oct. 1997.
- [9] Lennon, J., "Aspects of Large World Wide Web Systems", *Proceedings of WebNet International Conference*, 1996. Available on <http://aace.virginia.edu/aace/conf/webnet/html/220/220.htm>
- [10] Ignazio, J., *Introduction to Expert Systems: The Development and Implementation of Rule-Based Expert Systems*, McGraw-Hill, New York. USA, 1991.
- [11] Cloete, I. and Zurada, J., *Knowledge-Based Neurocomputing*, MIT Press, Massachusetts, USA, 2000.