PROCEEDINGS OF
THE 2007 INTERNATIONAL CONFERENCE ON
ARTIFICIAL INTELLIGENCE

ICAI 2007

Volume II

Editors

Hamid R. Arabnia
Mary Qu Yang, Jack Y. Yang

Associate Editors

Nikolaos G. Bourbakis, Halil Ceylan, Yuehui Chen
Georgios A. Demetriou, David Dodds, Laura L. Elnitski
Okan K. Ersoy, David de la Fuente, Joydeep Ghosh
Peter M. LaMonica, Guo-Zheng Li, Raymond A. Liuzzi
Jun Ni, Andrzej Niemierko, Jose A. Olivas, Bill Rand
Gene Simmons, Ashu M. G. Solo, Andrew Sung
Sule Yildirim, Yanqing Zhang

WORLDCOMP'07
June 25-28, 2007
Las Vegas Nevada, USA
www.world-academy-of-science.org

This set of volumes contain papers presented at The 2007 International Conference on Artificial Intelligence (ICAI'07). Their inclusion in this publication does not necessarily constitute endorsements by editors or by the publisher.

ISBN: 1-60132-023-X, 1-60132-024-8 (1-60132-025-6)
Printed in the United States of America

CSREA Press
U. S. A.

SESSION: SECOND ORDER META–PROGRAMMING – COGNITIVE FUSION AND AUTONOMOUS ROBOTS II

Fuzzy Term-Evaluation in an Information System to Manage Airport User Rights

Jesús Serrano-Guerrero	José A. Olivas	E. Herrera-Viedma	Javier de la Mata
Dept. of Information Technologies and Systems, Univ. of Castilla-La Mancha, Paseo de la Universidad 4, 13071-Ciudad Real, Spain, jesus.serrano@uclm.es	Dept. of Information Technologies and Systems, Univ. of Castilla-La Mancha, Paseo de la Universidad 4, 13071-Ciudad Real, Spain, joseangel.olivas@uclm.es	Dept. of Computer Science and Artificial Intelligence, University of Granada 18071-Granada, Spain viedma@decsai.ugr.es	Dept. of Information Technologies and Systems, Univ. of Castilla-La Mancha, Paseo de la Universidad 4, 13071-Ciudad Real, Spain, javier.mata@uclm.es

Abstract

In this paper, an approach for building small information systems for a well-known domain is presented. The domain chosen is "airport user rights and obligations". The main points of this architecture are a database of questions and answers and a fuzzy method to measure the relation between both parts (question-answer fuzzy matching).

Keywords: Frequently Asked Questions, Information Retrieval System, airport user rights, fuzzy evaluation.

1. Introduction

Nowadays, there are a lot of Information Retrieval Systems which are important tools in our daily life: search engines, meta-searchers [3], question answering systems, etc. All of them are based on two main factors: the query and the answer. A lot of these ones are based on complicated architectures, for example, based on agents, where a lot of parts of the architecture are focused in identifying the context of the required information. But when there is a well-known domain, the base of these systems could be easier, only based on the co-occurrence of words.

Classic document representation models (Boolean, probabilistic and vector space) in information retrieval [7] are used to model the conceptual load of a piece of information. Depending of the Information Retrieval System this piece of information will be different: a document, web pages, products, books, etc. These pieces of information have to include the system answers.

On the other hand, the query represents the information necessity of the user. The query formulation is very important to express exactly the user information needs. There are several kinds of query formulations: single terms or connected by logical operators, terms selected from menus, list of questions, questions written in natural language, etc. The kind of query depends on the own system design or the ability or experience of the user who is going to use the system.

It is not usual to know how to deeply use Information Retrieval Systems (IRS). User frequently do not know how to ask using only terms or terms combined with logic operators (AND, OR, etc.). There are a lot of small systems able to answer certain questions about a certain domain such as FAQ's (Frequent Asked Question), however, people usually would prefer to use alternative methods such as making a question in natural language, because it is more intuitive.

A lot of techniques to model each part of the IR systems can be used. Nowadays, Soft-computing (and mainly fuzzy logic [9]) based techniques play an important role and are one of the most interesting tools for modeling IR systems, for example, a lot of works are developed using fuzzy indexing [10].

All of these points have been taking into account to design the information system described below.

1.1 System Context

The main objective of the application that is being developed is to answer users' questions about airport problems in Spain. Then, one of the main tasks will be to summarize all the most recent laws about airport user rights and obligations for Spanish users, so the application is being developed for Spanish airports and language. These laws can be applied in Spain and in all the countries of the European Union and the Schengen Agreement countries. The 1985 Schengen Agreement is an agreement among European states which allows for the abolition of systematic border controls between the participating countries. It also includes provisions on common policy on the temporary entry of persons (including the Schengen Visa), the harmonization of external border controls and cross-border police co-operation.

There are a lot of moments in which passengers could have problems in the airport: baggage loss, flight delays, overbooking, etc., or simply airport users do not know which their rights are: possibility for transporting animals or guns into the aircraft, procedures to report wrong situations, etc.

Until now, information desks in some posters located in different terminals with brief and incomplete comments about the laws are the only references shown in Spanish airports. Furthermore, this information is written in languages that are not always perfectly understood by passengers because they usually are in international airports and the laws have a lot of complicated terms that only native speakers can understand. In conclusion, it is necessary a tool that is able to provide the requested information by airport users and so, the passengers won't have to read the Spanish or European laws to know their rights.

Perhaps a lot of airport users are used to navigate in Internet. Most of them bought their flights by means of e-commerce, but surely, there are other people that do not know the technologies related to Internet, older people for example. So, it is possible that this kind of people does not understand what a FAQ (Frequently Asked Questions) is, but they are able to write a query in natural language in a text box and wait for an answer.

In the following section, the main components of the system are described. In the third section, the focus is the necessary stages to obtain an answer in the system. The fourth section describes in detail the fuzzy evaluation query-answer. Finally, we conclude with some conclusions and a proposal of studies to be developed in the future.

2. System Components

The main component of the system is an obligations, rights and suggestions database. The obligations, rights and suggestions are relevant for airport users and most of them are included in the Spanish and European laws. System outputs are mainly focused in these obligations, rights and suggestions. There is a question set stored in the database apart from this information (possible answers). These are the questions the system is able to answer. If a user sends a query, the system proposes him a set of the most similar stored ones.

So, the two main logical elements of the system are: questions and answers. An expert has defined the stored questions which must be answered by the system and the answers for each question taking into account the laws. Each question and answer is defined using a term set selected by the expert and all the terms are stored in a database. All the stored terms compose the system corpus.

2.1 Questions

There is a question set which the system is able to answer. These questions are stored in the system database manually introduced by an expert.

Each question is composed by a term set that is part of the system corpus. These terms are also evaluated by an expert, assigning a fuzzy degree to each one. Each fuzzy degree expresses the importance of the term to the question. For instance:

How much can I claim due to losing baggage?

Can be expressed by means of the term set:

Money, lose, baggage

and finally with the fuzzy degrees, it is obtained a weighted term vector that represents the question:

	Lose	Baggage	Money
Question1	0.7	0.8	0.9

2.2 Answers

The system stores a lot of possible answers related to the different topics. All of them are based on the text of official Spanish and European laws. Each answer is described for several fields:

1. Topic: Each law deals with different topics: baggage loss, animal transporting, dangerous baggage…
2. Subtopic: Each topic can be divided into subtopics.
3. Definition: The answer extracted from the law.
4. Important terms: Terms that describe better what is the answer talking about.
5. Question related to the answer: Several questions that this answer could answer.
6. Associated laws: The identification of the law or laws where this answer can be found.

This information is shown to the user.

As well as the questions, each answer is composed by a term set that is part of the system corpus. These terms are also evaluated by an expert, assigning a fuzzy degree to each one. Each fuzzy degree expresses the importance of the term in the answer.

So, the relation between questions, answers and terms is shown in Fig.1..

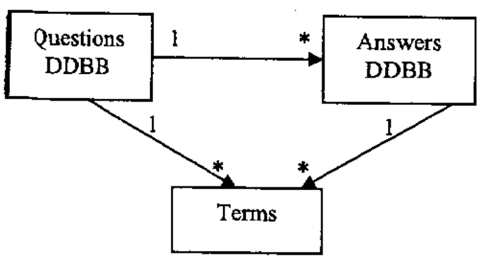

Fig.1. Database logical components.

3. Searching Process

The main objective of the system (Fig.2) is to present an interface where a user can write a question using the keyboard, and then a set of possible questions related to original one is returned with an unique answer for the user query. These alternative questions are taken from the database. If the answer provided to the user is not satisfactory, the user can select an alternative one. If the user does not select any alternative query (because he considers that alternative queries are not relevant), the system will not be able to give another answer to the original user question.

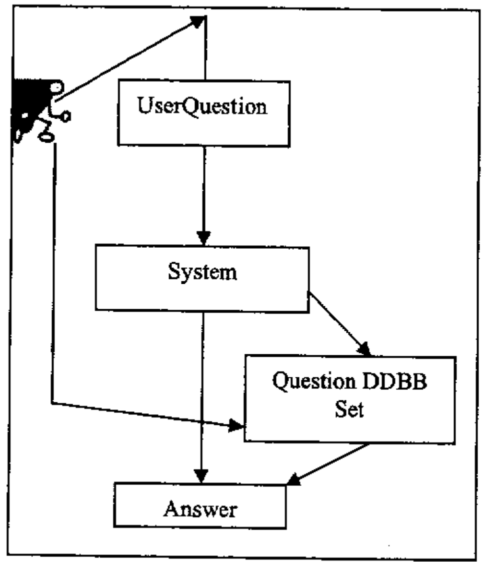

Fig.2. Main question-answering process

3.1 Stages

The main stages of the searching process are (Fig. 3):

1. Parsing.
2. Query Representation.
3. Fuzzy Term-Evaluation:
   3.1. Matching Query-Answer.
   3.2. Alternative question evaluation.
4. User evaluation.

1. Parsing:

   Natural language processing which simplifies the user query through the use of stoplist and Stemming algorithms [6].

2. Query Representation:

   The terms from the previous stage and new related ones added by the Spanish version of the electronic dictionary WordNet[1] are used to create a term vector that represents the user query, in a similar way as in authors previous works [2, 4].

3. Fuzzy Term-Evaluation

   This stage selects an answer for the user query and returns a set of alternative queries from the database.

   3.1. Matching Question-Answer:

[1] http://www.lsi.upc.edu/~nlp/web/index.php (last visit 24-04-2007)

This is the most complex stage. It will be explained in the next section.

3.2. Alternative questions evaluation:
The system retrieves several questions from the database by means of syntactic matching between the terms of the user query and the queries stored in the database.

4. User evaluation:

At the end of the process, if the user selects a query from the alternatives returned by the system, the original user query can be added to the query database and associated to the answer of the query selected by the user.

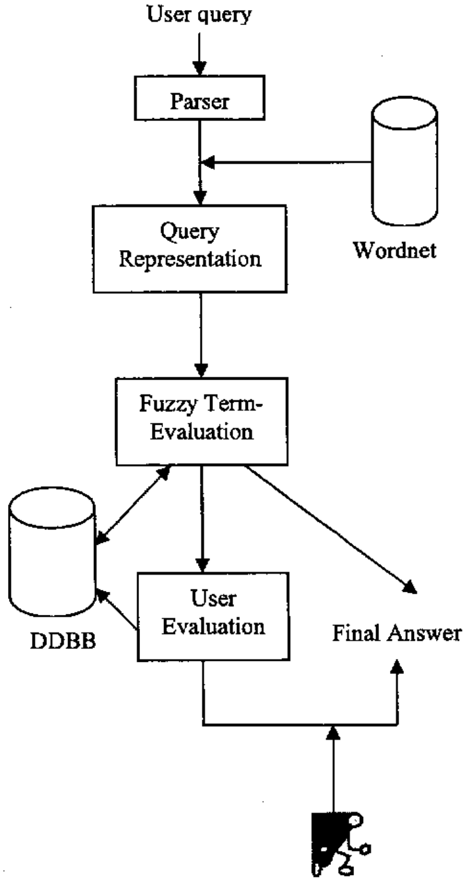

Fig.3. Stages

5. Fuzzy Term-Evaluation

The evaluation process is the most interesting of the above mentioned stages. It is explained with the following example.

If there are two possible codified questions that the system is able to answer:

Question1:
Which are the necessary procedures when I lose my luggage?

Question2:
How much is it paid when baggage is lost?

Each question can be represented by a term set:

Question1: lose, baggage, procedures.
Question2: lose, baggage, money.

Only with these terms, system corpus would be:

Corpus: {lose, baggage, money, procedure}

Now, the expert can assign a fuzzy weight for each term taking into account the relevance of the term to the question. For example, the previous questions can be evaluated as follows:

	Lose	Baggage	Money	Procedure
Question1	0.7	0.7	0.2	0.7
Question2	0.7	0.8	0.7	0.1

Each answer is evaluated by the expert too. So, two possible answers for each question will be composed by the same or related terms: For instance, two possible answers codified are:

	Answer1	Answer2
Lose	0.7	0.7
Baggage	0.7	0.7
Money	0.8	0.2
Procedure	0.2	0.8

So, the database is composed by two main sets:

- Query set.
- Answer set.

Each one is represented as a relation R defined on the Cartesian product A x B of the set A={queries/answers} and B={corpus terms}:

$$R = \int_{axb} \mu_R(a,b) / (a,b)$$

where the membership function of R

$$\mu_R = A x B \rightarrow [0,1]$$

is defined by the expert.

To evaluate which is the most relevant answer for the original user query it is necessary to use the composition operation (max-min) to calculate the fuzzy relation degrees among the queries and the answers.

The most generic words are deleted because they are not relevant when the context is detected. In this application the context is relatively easy to detect because is a small application and each answer is stored with its topic and subtopic. If the more generic words are not deleted, it would not be possible to distinguish between the first answer and the second answer when the system applies the composition operation:

	Answer1	Answer2
Question1	0.7	0.7
Question2	0.7	0.7

So, if the topic or subtopic of the answers is "baggage lost", the terms related to the context, such as "baggage" or "lose", must be deleted. Now, the new question-term and term-answer relations are:

	Money	Procedure
Question1	0.2	0.7
Question2	0.7	0.1

	Answer1	Answer2
Money	0.8	0.2
Procedure	0.2	0.8

And the result of the application of the max-min operator is:

	Answer1	Answer2
Question1	0.2	0.7
Question2	0.7	0.2

Then, an answer is closer to a question than the other one.

The system is prepared to answer the question set, but it is possible that there were user questions that can not be answered or questions more or less general than the original one, for example:

Which are the necessary procedures to earn money due to losing baggage?

This question can represent the two previous ones, so it would be interesting that the system could answer with both answers.

If the user labels the answer as correct, the system would add the new question to its question database, and the new question would be composed by the previous terms with the new fuzzy weights calculated using a fuzzy union operation (T-conorm: in this case MAX) between both questions.

	Lose	Baggage	Money	Procedure
Question1	0.7	0.7	0.2	0.7
Question2	0.7	0.8	0.7	0.1
NewQuestion	0.7	0.8	0.7	0.7

If the user labels the answer as correct, the system would add a new item of the questions set. The new question is composed by the previous terms with the new fuzzy weights calculated using the above T-conorm between both answers:

	Lose	Baggage	Money	Procedure
Answer1	0.7	0.7	0.2	0.7
Answer2	0.7	0.8	0.7	0.1
NewAnswer	0.7	0.8	0.7	0.7

And the answers continue being the same ones but the new weights in the fuzzy question-answer relation indicates there are several answers that can be retrieved by the system as a good solution.

	Answer1	Answer 2
Lose	0.7	0.7
Baggage	0.7	0.7
Money	0.8	0.2
Procedure	0.2	0.8

Deleting the more generic terms:

	Money	Procedure
Question1	0.2	0.7
Question2	0.7	0.1
NewQuestion	0.7	0.7

	Answer1	Answer2
Money	0.8	0.2
Procedure	0.2	0.8

And the fuzzy logic composition shows:

	Question1	Question 2	NewQuestion
Answer1	0.2	0.7	0.7
Answer2	0.7	0.2	0.7

The new question is closed to the first and the second answer, and then the application would return the two answers.

6. Conclusions and future work

The application proposed is a simple information system for Spanish language to solve problems to airport users. It is only proposed a first approach to evaluate the answer-question relation. The system represents a possible solution for small information retrieval systems.

One of the main problems of the system is to determine the quantity of information that can be provided by an answer: a number, a phrase or a complete paragraph. It is necessary to redefine the answers to develop a more precise system close to the question-answering systems proposes by Zadeh [8].

The fuzzy operators used for modeling the process are very basic, so it is necessary to test different ones to improve the system results.

The system could also add the new user queries to the system and automatically associates fuzzy weights to the terms.

Nowadays the application is only for Spanish language but could be modified to be used by English speakers using dictionaries such as WordNet [5] and syntactic analyzers such as GATE [1].

Acknowledgments

Partially supported by SCAIWEB project, JCCM, Spain and Fom2005-01 project, Ministerio de Fomento, Spain.

References

[1] GATE (General Architecture for Text Engineering) http://www.gate.ac.uk/

[2] J. Gonzalo, F. Verdejo, I. Chugur, J. Cigarrán, "Indexing with WordNet synsets can improve text retrieval," Proceedings of the COLING/ACL'98 Workshop on Usage of WordNet for NLP, Montreal, 1998, pp. 38-44.

[3] P. Garcés; J. A Olivas; F. P. Romero,: "Concept-matching IR systems versus Word-matching IR systems: Considering fuzzy interrelations for indexing web pages". Journal of the American Society for Information Science and Technology JASIST, 57 (4): 564-576, 2006.

[4] R. Mihalcea, D. Moldovan, "Semantic Indexing using WordNet Senses," Proceedings of ACL Workshop on IR & NLP, Honk Kong, 2000.

[5] WORDNET. A lexical database for the English language. http://www.cogsci.princeton.edu/~wn/

[6] M. F. Porter, "An Algorithm for Suffix Stripping". Program. 1980; 14(3): 130-137.

[7] R. Baeza-Yates, B. Ribeiro-Neto: Modern Information Retrieval. Addison Wesley. 1999.

[8] L. Zadeh: "From search engines to Question-Answering System: The need for new tools", Proceedings of the First International Atlantic Web Intelligence Conference AWIC 2003, Madrid.

[9] L. Zadeh, "Fuzzy sets and systems", in Systems and Theory (J.Fox), Polytechnic Press, New York, 1965, pp. 29-37.

[10] M. Nikravesh. "Fuzzy conceptual-based search engine using conceptual semantic indexing", Proceedings of the 2002 NAFIPS Annual Meeting, 2002, pp. 146–151.