A computer-supported learning system to help teachers to teach Fuzzy Information Retrieval Systems

E. Herrera-Viedma · A. G. López-Herrera · S. Alonso ·
J. M. Moreno · F. J. Cabrerizo · C. Porcel

Abstract This paper describes a computer-supported learning system to teach students the principles and concepts of Fuzzy Information Retrieval Systems based on weighted queries. This tool is used to support the teacher's activity in the degree course *Information Retrieval Systems Based on Artificial Intelligence* at the Faculty of Library and Information Sciences at the University of Granada. Learning of languages of weighted queries in Fuzzy Information Retrieval Systems is complex because it is very difficult to understand the different semantics that could be associated to the weights of queries together with their respective strategies of query evaluation. We have developed and implemented this computer-supported education system because it allows to support the teacher's activity in the classroom to teach the use of weighted queries in FIRSs and it helps students to develop

E. Herrera-Viedma · A. G. López-Herrera (✉)
Department of Computer Sciences and A.I., University of Granada, 18071 Granada, Spain
e-mail: lopez-herrera@decsai.ugr.es

E. Herrera-Viedma
e-mail: viedma@decsai.ugr.es

S. Alonso
Department of Software Engineering, University of Granada, 18071 Granada, Spain
e-mail: zerjioi@ugr.es

J. M. Moreno
Department of Information and Communication Engineering, University of Murcia,
30071 Murcia, Spain
e-mail: jmmoreno@um.es

F. J. Cabrerizo
Department of Software Engineering and Computer Systems, Distance Learning
University of Spain (UNED), 28040 Madrid, Spain
e-mail: cabrerizo@issi.uned.es

C. Porcel
Department of Computer Sciences, University of Jaén, 23071 Jaén, Spain
e-mail: cporcel@ujaen.es

self-learning processes on the use of such queries. We have evaluated the performance of its use in the learning process according to the students' perceptions and their results obtained in the course's exams. We have observed that using this software tool the students learn better the management of the weighted query languages and then their performance in the exams is improved.

1 Introduction

Due to the growth of e-business, the Web has become a critical part of many real-world systems. Thus, it is increasingly important that information technology professionals and students are proficient and knowledgeable in various Web technologies like (Baeza-Yates and Ribeiro-Neto 1999) Web mining, query processing, Information Retrieval (IR) models, search engines, meta-search engines, recommender systems, information filtering, Web quality evaluation, etc., which are also evolving at a fast rate, making it critical to keep up-to-date with them (Chau et al. 2003).

At the Faculty of Library and Information Science at the University of Granada there are different degree courses that address these evolving needs. In particular, there exists a degree course called "Information Retrieval Systems based on Artificial Intelligence" which deals with the study and analysis of Artificial Intelligence tools applied in the design of Information Retrieval Systems (IRSs). The key goals of this course are to learn the foundations of Fuzzy Logic tools and Genetic Algorithms and their application in the design of IRSs. As it is known, both are important Soft Computing tools (Bonissone 1997) and are being satisfactorily applied in the development of Web access technologies (Crestani and Pasi 2000; Herrera-Viedma et al. 2006; Nikravesh et al. 2002).

Fuzzy IRSs (FIRSs) are those IRSs that use the potential of the fuzzy tools to improve the retrieval activities (Crestani and Pasi 2000; Herrera-Viedma et al. 2006). The teaching of FIRSs in the degree course "Information Retrieval Systems based on Artificial Intelligence" is focused on models of FIRSs that use weighted queries to improve the representation of user information needs and fuzzy connectives to evaluate such queries. In this course, blackboard classroom exercises for teaching and practising the use of weighted queries and fuzzy connectives are used. However, in our teaching experience we observed that this is not enough to show learners the searching skills of FIRSs because it is very difficult to visualize the different semantics that could be associated to the weights of queries together with their respective strategies of query evaluation in a blackboard.

Researchers have found that using computer-supported learning systems in instructional contexts may provide students with opportunities to promote their understanding of phenomena in science and to facilitate the visualization of abstract and unobservable concepts (Alessi and Trollip 1991; Hegarty 2004; Stratford 1997). Computer-based instruction software allows students to develop self-learning processes which offers them more flexible learning opportunities, independent from time and place, and to learn at their own pace (Eteokleous 2008).

IR instruction is an obvious application for computer-supported learning systems. The advantage of using computer-supported learning systems is that the learner gets a realistic

feeling of the particular IRS used and he can develop self-learning processes on typical operations of IRSs (Halttunen and Sormunen 2000). To do that, it is possible to use real world search engines like Google, Altavista, Lycos, etc., or to build ad-hoc training IRSs (Caruso 1981; Chau et al. 2003; Griffith and Norton 1981; Halttunen and Sormunen 2000; Markey and Atherton 1978).

There are very few training IRSs (Halttunen and Sormunen 2000) and, particularly, a fuzzy IR training system does not exist. Furthermore, existing IR training systems present several shortcomings (Halttunen and Sormunen 2000):

(1) They do not give feedback about the performance or success of user queries,
(2) it is not possible to observe or visualize how a user query is evaluated, and
(3) it is not possible to compare the performance of different types of user queries and different evaluation procedures of user queries.

The main aim of the paper is to introduce a computer-supported learning system to improve the teaching of FIRSs at the degree course "Information Retrieval Systems based on Artificial Intelligence" at the Faculty of Library and Information Sciences at the University of Granada, which overcomes the above shortcomings and, particularly, the teaching problems of FIRS. The system was used as first time at the academic year 2006–2007. It provides an environment for demonstrating the use and performance of weighted queries with different semantics and their evaluations using different fuzzy connectives. It offers students the opportunity to see and compare the achieved results of different weighted queries. Student can choose (i) different semantics (threshold, relative importance, ideal importance, quantitative) (Herrera-Viedma 2001b; Herrera-Viedma et al. 2005) to formulate weighted queries, (ii) different fuzzy connectives to evaluate such queries (maximum, minimum, Ordered Weighted Averaging (OWA) operators, Induced OWA operators, Linguistic OWA operators, and Linguistic Weighted Averaging operators) (Chiclana et al. 2004; Chiclana et al. 2007; Herrera and Herrera-Viedma 1997; Herrera et al. 1996; Yager 1987, 1988; Yager and Filev 1999), and (iii) different expression domains (numerical and fuzzy ordinal linguistic one) (Herrera and Herrera-Viedma 2000; Herrera-Viedma 2001b; Herrera-Viedma et al. 2007) to assess weights associated with queries. Furthermore, several standard test collections (ADI, CISI, CRANSFIELD, TREC, etc.) can be used. The system presents visualization tools to better show evaluation processes of user queries. We have to point out that this system is just an educational tool that can be employed to help teachers activities, and not an independent teacher itself. It allows students to develop self-learning processes, and, as it is known, this is an important motivational factor supporting the learning process which leads to increase learning gains (Yaman et al. 1784). Additionally, we evaluate the performance of its use in the learning process according to the students' perceptions and their results obtained in the course's exams. We observe that with this teaching tool the students learn better the complex skills that those FIRS provide, their motivation is increased, and their performance in exams is improved.

The paper is structured as follows. In Sect. 2 we describe the educational context in which we use our computer-supported learning system for FIRSs, the models of FIRSs that the system contains, and the problems that their teaching arises. Section 3 describes the structure and performance of the learning system. In Sect. 4, we evaluate the system, discuss some lessons learned from our experience and suggest some possible uses and improvements of our computerized system. And finally, in Sect. 5 some conclusions are pointed out.

2 Preliminaries

In this section, we present the context of our experience, the basic notions of the models of FIRSs that the learning system contains, and some problems that we have detected to teach FIRSs in blackboard classes.

2.1 Educational context

The Department of Computer Science and Artificial Intelligence (DECSAI) at the University of Granada teaches IRSs based on Artificial Intelligent at both graduate and postgraduate level at the Faculty of Library and Information Sciences. This paper reports on the experiences of teaching IRSs at the graduate level.

In the graduate level we teach a course "Information Retrieval Systems based on Artificial Intelligence". As aforementioned, this course deals with the study and analysis of Fuzzy Logic tools and Genetic Algorithms applied in the design of Information Retrieval Systems (IRSs). The teaching of FIRSs is focused on those FIRSs that allow to use weighted queries to represent user information needs. This is a fifth year course (fifth year is the final year for graduates in Spain) and so, it is a research-led course. This means that the course should be built on students' existing knowledge of Library and Information Sciences, gained from their previous four years of study, but should also present material that is informed by the latest research ideas. This implies that the students appreciate the open problems in IR, learn about new approaches and more radical solutions that are still in the laboratory stages of development. In this course we are particularly keen that students learn about the breadth of FIRSs problems and domains as much of their current experience is with Web search engines.

Some important characteristics of our educational framework are the following:

- *An adequate size of the class* with respect to the number students, we should point out that from the academic year 2004–2005 we have a reduced cohort of students, between 15 and 20 in every group.
- *Teaching procedure* the main instruction method is lecturing supplemented with relatively large sized tutorials and laboratory sessions which gave practical experience of IR packages, as the one presented in this paper.
- *Adequate technological formation of the students* at the Faculty of Library and Information Sciences the students receive a solid formation in Information Technologies and this helps us to develop the teaching of different models of FIRSs.

2.2 Models of FIRSs

As aforementioned, our computer-supported learning system allows us to teach different models of FIRSs which are based on weighted user queries. These models of FIRSs present the following components: a documentary archive, a query system and a query evaluation procedure.

2.2.1 Documentary archive

We assume a documentary archive built like in an usual IRS (Baeza-Yates and Ribeiro-Neto 1999; Salton and McGill 1983). The archive stores a finite set of documents $\mathcal{D} = \{d_1, \ldots, d_m\}$, a finite set of index terms $\mathcal{T} = \{t_1, \ldots, t_l\}$, and the representation R_{d_j} of each

document d_j characterized by a numeric indexing function $\mathcal{F}: \mathcal{D} \times \mathcal{T} \rightarrow [0,1]$ which assigns a numeric weight to each index term t_i. In fuzzy notation, R_{d_j} is a fuzzy set represented as:

$$R_{d_j} = \sum_{i=1}^{l} \mathcal{F}(d_j, t_i)/t_i$$

where $\mathcal{F}$ represents the significance of an index term in describing the content of a document. $\mathcal{F}(d_j, t_i) = 0$ implies that the document d_j is not at all related to the concept(s) represented by index term t_i and $\mathcal{F}(d_j, t_i) = 1$ implies that the document d_j is perfectly represented by the concept(s) indicated by t_i. In standard test collections $\mathcal{F}$ is obtained using a $tf \cdot idf$ scheme (Salton and McGill 1983).

2.2.2 Query system

The implemented FIRSs present a query system based on a weighted Boolean query language. Each user query is expressed as a combination of the weighted terms which are connected by the logical operators AND ($\wedge$), OR ($\vee$), and NOT ($\neg$). The weights associated with the query terms could be numerical values assessed in [0, 1] or linguistic values taken from a linguistic term set $\mathcal{S}$ defined in a fuzzy ordinal linguistic context (Herrera and Herrera-Viedma 2000; Herrera-Viedma 2001a, b; Herrera-Viedma et al. 2005, 2007) (see Appendix 1).

In this context, a user query is any legitimate Boolean expression whose atomic components (atoms) are pairs $<t_i, w_i>$, $t_i \in \mathcal{T}$ and being $w_i \in \mathcal{I}, \mathcal{I} \in [0,1]$ or $\mathcal{I} \in \mathcal{S}$ the weight associated to the term t_i by the user. Then, the set $\mathcal{Q}$ of the legitimate weighted Boolean queries is defined by the following syntactic rules:

(1)   Atomic queries: $\forall q = \; <t_i, w_i> \; \in \mathcal{T} \times \mathcal{I} \Rightarrow q \in \mathcal{Q}$.
(2)   Conjunctive queries: $\forall q, p \in \mathcal{Q} \Rightarrow q \wedge p \in \mathcal{Q}$.
(3)   Disjunctive queries: $\forall q, p \in \mathcal{Q} \Rightarrow q \vee p \in \mathcal{Q}$.
(4)   Negated queries: $\forall q \in \mathcal{Q} \Rightarrow \neg(q) \in \mathcal{Q}$.
(5)   All legitimate queries $q \in \mathcal{Q}$ are only those obtained by applying rules 1–4, inclusive.

By assigning weights in queries, users specify restrictions on the relevant documents to retrieve. In the literature we find four kinds of semantics to interpret the weights in queries (Herrera-Viedma 2001b; Kraft et al. 1994):

- *Relative importance semantics* this semantics defines query weights as measures of the relative importance of each term with respect to the others in the query. By associating relative importance weights to terms in a query, the user is asking to see all documents whose content represents the concept that is more associated with the most important term than with the less important ones. In practice, this means that the user requires that the computation of the relevance degree of a document should be dominated by the more heavily weighted terms.
- *Threshold semantics* this semantics defines query weights as satisfaction requirements for each term of the query to be considered when matching document representations to the query. By associating threshold weights with terms in a query, the user is asking to see all the documents sufficiently related to the topics represented by such terms. In practice, this means that the user will reward a document whose index term weights $\mathcal{F}$ exceed the established thresholds with a high relevance degree.

- *Perfection semantics* this semantics defines query weights as descriptions of ideal or perfect documents desired by the user. By associating weights with terms in a query, the user is asking to see all the documents whose content satisfies or is more or less close to his ideal information needs as represented in the weighted query. In practice, this means that the user will reward a document whose index term weights are equal to or at least near to term weights for a query with the highest relevance degrees.
- *Quantitative semantics* this semantics defines query weights to express criteria that affect the quantity of the documents to be retrieved, i.e., constraints to be satisfied by the number of documents to be retrieved.

2.2.3 Query evaluation procedure

The evaluation procedure of weighted queries acts as a constructive bottom-up process that includes two steps:

- Firstly, the documents are evaluated according to their relevance only to atoms of the query. Then, a partial relevance degree is assigned to each document with respect to each atom in the query.
- Secondly, the documents are evaluated according to their relevance to Boolean combinations of atomic components (their partial relevance degree), and so on, working in a bottom-up fashion until the whole query is processed. In this step, a total relevance degree is assigned to each document with respect to the whole query.

We represent the query evaluation procedure using an evaluation function $\mathcal{E}: \mathcal{D} \times \mathcal{Q} \to \mathcal{I}$. Depending on the kind of query, $\mathcal{E}$ obtains the relevance degree RSV_j of any $d_j \in \mathcal{D}$ according to the following rules:

(1) Evaluation of an atomic query:

$$\mathcal{E}(d_j, <t_i, w_i>) = g^1(\mathcal{F}(d_j, t_i), w_i) = RSV_j,$$

where g^1 is a matching function defined according to the semantics associated to w_i. The four kind of semantics with different interpretations or matching functions have been considered in our computer-supported learning system for FIRSs, i.e., classical threshold semantics (Buell and Kraft 1981; Waller and Kraft 1979), symmetrical threshold semantics (Herrera-Viedma 2001b), improved threshold semantics (Herrera-Viedma et al. 2005), relative importance semantics (Herrera-Viedma et al. 2005; Yager 1987), improved relative importance semantics (Herrera-Viedma 2001a; Herrera-Viedma et al. 2003), classical perfection semantics (Bordogna and Pasi 1993), non-symmetrical perfection semantics (Kraft et al. 1994), quantitative semantics (Herrera-Viedma 2001b).

(2) Evaluation of a conjunctive query:

$$\mathcal{E}(d_j, q \wedge p) = \mathcal{E}(d_j, q) \overset{FC}{\bigwedge} \mathcal{E}(d_j, p),$$

where $\bigvee^{FC}$ is a fuzzy connective that models a combination of values similar to a t-norm.

(3) Evaluation of disjunctive query:

$$\mathcal{E}(d_j, q \vee p) = \mathcal{E}(d_j, q) \overset{FC}{\bigvee} \mathcal{E}(d_j, p),$$

where $\bigvee^{FC}$ is a fuzzy connective that models a combination of values similar to a t-conorm.

(4) Evaluation of a negated query:

$$\mathcal{E}(d_j, \neg q) = \mathcal{N}eg(\mathcal{E}(d_j, q)),$$

where $\mathcal{N}eg$ is a complement operator of fuzzy sets.

We should point out that the fuzzy connectives that are applied in the evaluation procedure are the following: OWA operators (Yager 1988), Induced OWA operators (Chiclana et al. 2004, 2007; Yager and Filev 1999), weighted aggregations MAX and MIN (Yager 1987), Linguistic OWA operators (Herrera et al. 1996), and Linguistic Weighted Averaging operators (Herrera and Herrera-Viedma 1997).

2.3 Teaching problems for FIRSs

The main difficulties to teach FIRSs that we have detected during different academic years (we began to teach FIRSs in the academic year 1995–1996) are the following:

(1) *How to explain students the different interpretations of a kind of semantics?* For example, for threshold semantics we can use three different threshold proposals: classical threshold semantics (Buell and Kraft 1981; Waller and Kraft 1979), symmetrical threshold semantics (Herrera-Viedma 2001b) and improved threshold semantics (Herrera-Viedma et al. 2005). To understand their different meanings we have observed that the students need to process many examples and compare the results continuously, and it is very difficult to get it in blackboard classes. A similar situation happens when teching the perfection semantics and the relative importance semantics.

(2) *How to explain students the contradictions existing between different semantics?* For example, as it is demonstrated in (Herrera-Viedma 2001b) the threshold semantics and the perfection semantics are contradictory for values of index weight function $\mathcal{F}$ over the considered threshold value. To fully understand it, it is necessary to develop multiple examples and comparisons, and as in the previous problem, this is very difficult to achieve in blackboard classes.

(3) *How to explain students the models of FIRSs that propose to use the different semantics simultaneously in sthe ame query?* For example, in (Herrera-Viedma 2001a, b; Herrera-Viedma et al. 2003; Herrera-Viedma and López-Herrera 2007) different models of FIRSs simultaneously using different semantics are proposed. Students have problems to understand the formulation process and the evaluation process of such queries.

(4) *How to explain students the bottom-up evaluation procedure for weighted Boolean queries?* To help students to understand this evaluation procedure it would be convenient to draw in the blackboard the evaluation tree of each user query and to show the results in each step of evaluation, and this is not possible due to the limited space of a blackboard.

(5) *How to explain students the problems of the usual connectives, t-norm Min and t-conorm Max, used to model the logical connectives $\wedge$ and $\vee$, respectively?* The t-norm Min and t-conorm Max are usually applied in FIRSs to model the logical connectives $\wedge$ and $\vee$ in the evaluation of user queries. Therefore, they present the same problems than the intersection and union operators applied in classical Boolean

IRSs, i.e., restrictive and inclusive behaviour (Herrera-Viedma and López-Herrera 2007), respectively. To help students to understand this problem it is necessary to develop multiple exercises and compare the results, and this is not easy in a usual classroom environment.

(6) *How to explain students the use of the OWA and Linguistis OWA aggregation operators to model the logical connectives ∧ and ∨ in the evaluation procedure?* As it is known, OWA operators are "or-and" operators (Herrera et al. 1996; Yager 1988) and their behaviour can be controlled by means of the a weighting vector, which can be determined by means of an *orness* measure. To explain how to compute this weighting vector from the orness measure and how this weighting vector acts on the evaluation of user queries requires many practical exercises.

(7) *How to explain students the use of the weighted aggregation operators to model the relative importance semantics in the evaluation of user queries?* As it is known, the relative importance semantics is modelled by means of weighted aggregation operators (Herrera-Viedma 2001a; Yager 1987), and then the matching function used to model its interpretation in the evaluation of weighted queries depending on the aggregation operator used to model the logical connectives ∧ and ∨ in the evaluation of user queries. As above, to explain how to evaluate this semantics in user queries requires many practical exercises and this is difficult to develop in a blackboard class.

These are the main teaching problems that we propose to solve with our computer-supported learning system of FIRSs, which is introduced in the following section.

3 A computer-supported learning system to teach FIRSs

The computer-supported learning system of FIRSs that we present in this section has been designed to help us to overcome the problems mentioned in the above section. It also is useful to overcome the problems of existing IR training systems mentioned at the beginning. This system has been developed at the Faculty of Information and Library Science at the University of Granada as a useful training tool of FIRSs based on weighted queries (see http://sci2s.ugr.es/secabalab/software/FIRS-trainer/).

The goal of this software application is, on the one hand, to provide an environment for demonstrating students the performance of weighted queries and, in such a way, to aid them in their learning. On the other hand, to support the teacher's activity in the classroom to teach the use of weighted queries.

This learning tool is a Web-based application that is implemented in the Java programming language. It is composed of three main modules: (i) definition module of test collections, (ii) formulation module of weighted queries, and (iii) a visual execution module of weighted queries. We analyze all of them in detail in the following subsections.

3.1 Definition module of test collection

An experimental test collection consists of an archive, a collection of queries and relevance assessments indicating which documents are relevant in respect to a given query. Usually, the performance of a system is measured by means of the precision and recall achieved across the whole set of queries. As in (Halttunen and Sormunen 2000; Hull 1996) our goal is to encourage the analysis of the individuals queries, and therefore, we only need an
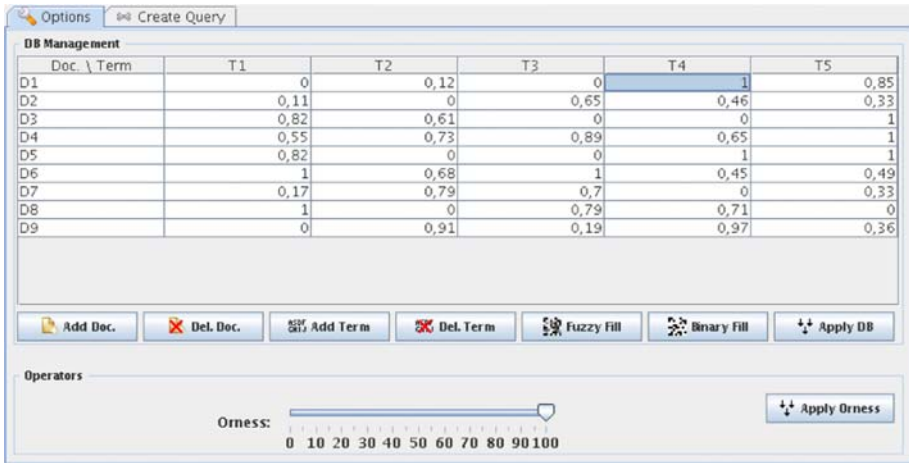
Fig. 1 Defining a test collection

instructional test collection. However, the tool also provides some standard test collections (ADI, CISI, CRANSFIELD, etc).

Students have the possibility of building their own test collections (see Fig. 1), i.e., toy test collections, to analyze the performance of the different weighted queries in FIRSs. In the definition of the test collection they can establish particular queries and the relevant documents of the archive of such queries.

3.2 Formulation module of weighted queries

The formulation module of weighted queries allows students to define their weighted queries (see Fig. 2). To define a weighted query they have to choose:
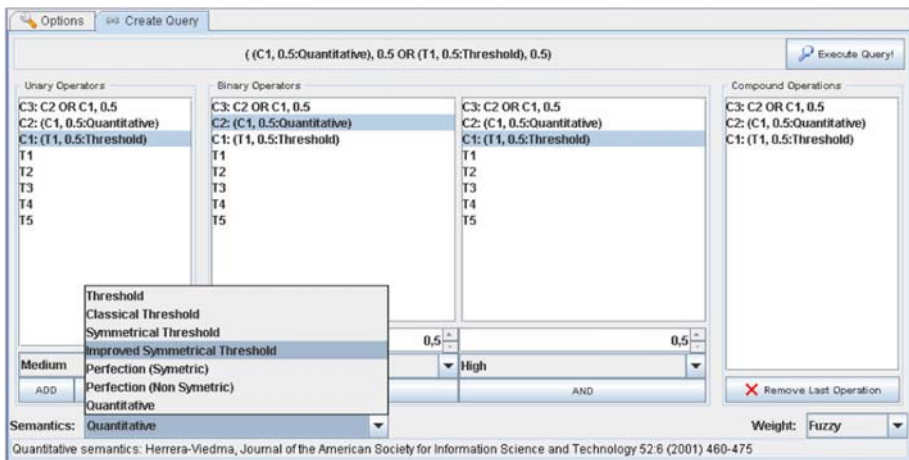

Fig. 2 Defining a weighted query

(1)   search terms,
(2)   Boolean connectives (AND, OR, and NOT),
(3)   query structure,
(4)   expression domain of weights (numerical or linguistic),
(5)   semantics to associate with the weights, and
(6)   values of weights.

3.3 Visual execution module of weighted queries

The execution module allows measuring and visualizing the performance of any weighted query. Before to execute any weighted query, students have to choose the fuzzy connectives that must be associated with the Boolean connectives of the weighted queries. Given that our system uses OWA operators, this can be done by choosing a level of *orness* (Yager 1988).

This module provides students a feedback on the evaluation of weighted queries by means of visual tools. This feedback is given by showing internal aspects of evaluations of weighted queries using evaluation trees (see Fig. 3). Furthermore, the module allows the visual comparison of the evaluation of different weighted queries (see Fig. 4).

3.4 Learning step by step: an example of use

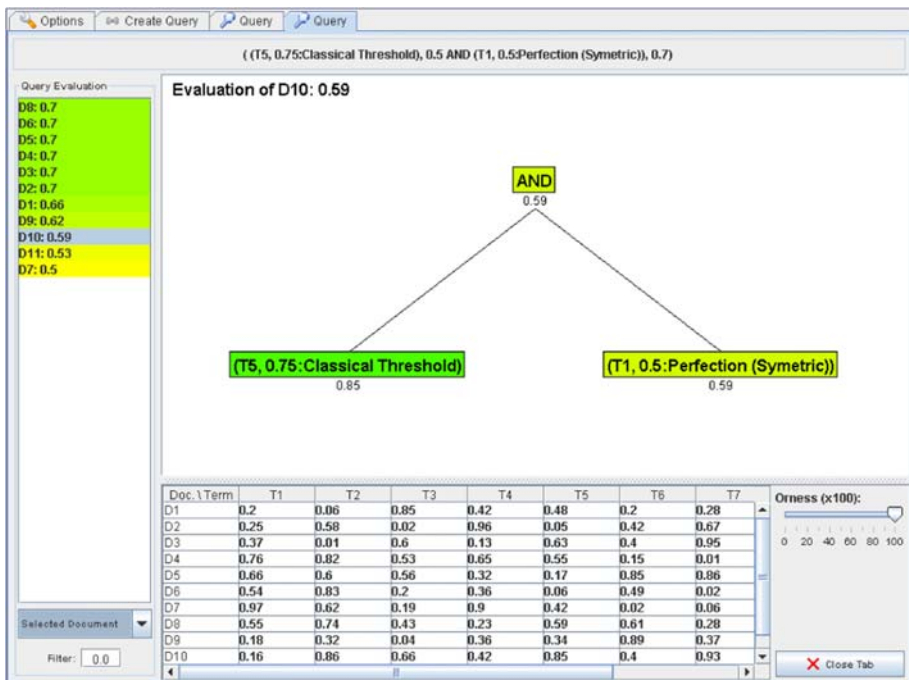In this subsection a learning session based on this software tool is briefly described.
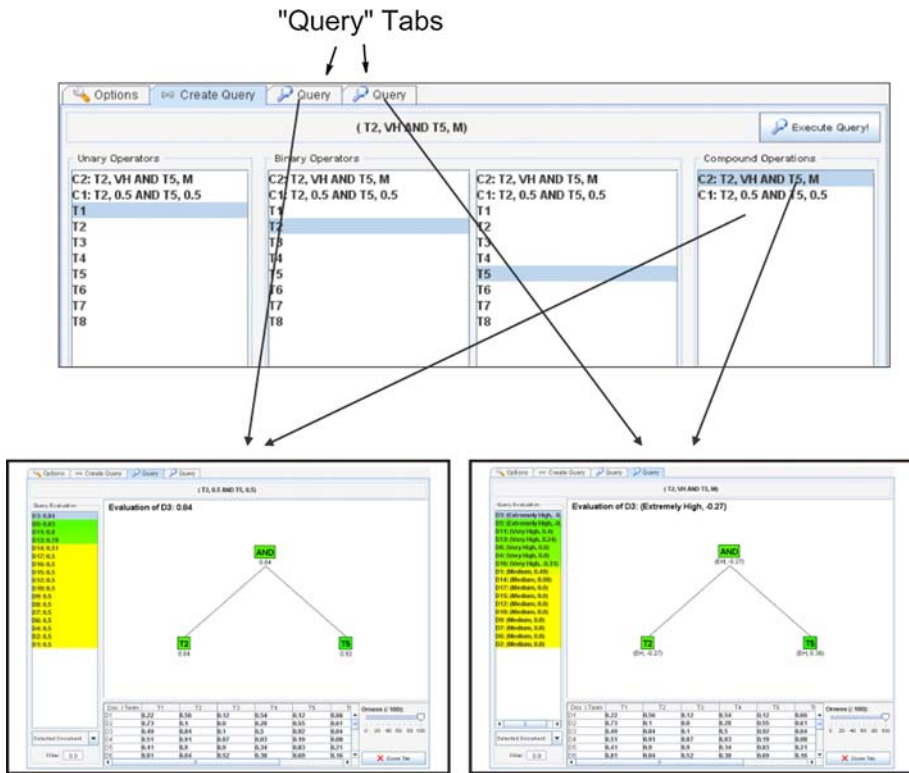

Fig. 3 Evaluation tree

Fig. 4 Visual comparison of the evaluation of different weighted queries

Let us suppose a student which wants to learn a fuzzy weighted IR model based on fuzzy ordinal linguistic modeling (see Appendix 1). Firstly, he should establish a test documentary archive. There are two possibilities to do it: i) defining it by hand, or ii) by using a standard test collection. To simplify, we suppose that this student chooses the first possibility.

Then, we suppose that he defines a small documentary archive containing a set of 17 documents $\mathcal{D} = \{d_1, \ldots, d_{17}\}$, represented by means of a set of 12 index terms $\mathcal{T} = \{t_1, \ldots, t_{12}\}$. Documents are indexed by means of a random numeric indexing function. The result of this process is shown in Fig. 5. Using the set of nine labels as that defined in Appendix 1, suppose that the student formulates a linguistic weighted query using two semantics simultaneously, threshold and quantitative ones:

$$q = < <t_1, M : ClassicalThreshold > , EH > \wedge$$

$$< <t_5, EH : Perfection(Symmetrical) > , L > .$$

In this query, two sub-queries $q_1 = <<t_5, M: \ Classical \ Threshold>, EH>$ and $q_2 = <<t_5, EH: \ Perfection \ (Symmetrical)>, L>$ are connected by the conjunctive connective AND ($\wedge$). With such a query, the user demands:

(1) q_1: The student is looking for almost all documents with a weight $\mathcal{F}$ for t_1 as high as possible[1], with a minimum value equal to the linguistic weight M. The quantitative

---

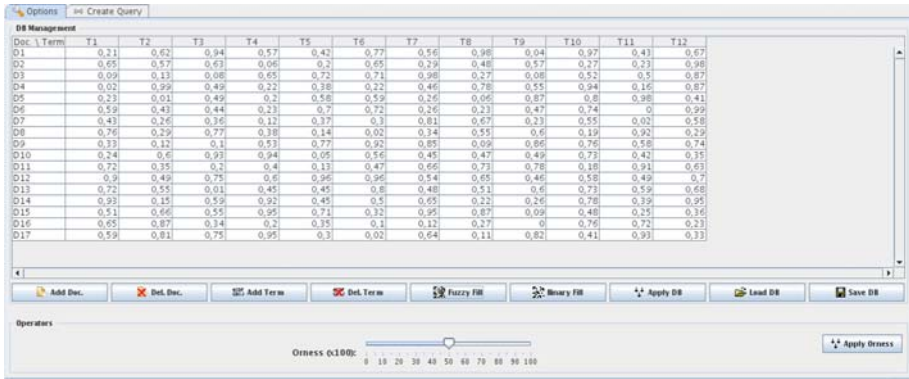[1] To do so, a classical threshold semantics is used.

Fig. 5 Defining a test collection

semantics represented by the weight equal to *EH* indicates that many documents should be retrieved for q_1.

(2) q_2: The student is looking for documents with a weight $\mathcal{F}$ for t_5 as high as possible, with a minimum value equal to the linguistic weight *EH*. Those documents with a very distant value of $\mathcal{F}$ to EH will be rejected. Additionally, only a small portion of relevant documents for t_5 will be taken into account. For this last, a quantitative restriction is imposed by the linguistic weight *L*.

(3) Only those documents that fulfill with the demands 1 and 2 will be considered as relevant.

To do so, the query *q* is built in subsequent steps as it is shown in Fig. 6.

Finally, the results of evaluating *q* for all relevant documents are shown. For example, Fig. 7, draws an evaluation tree with results of evaluating *q* for the document d_{14}. It can be seen how partial results for q_1 and q_2 are combined using a MIN t-norm to model the
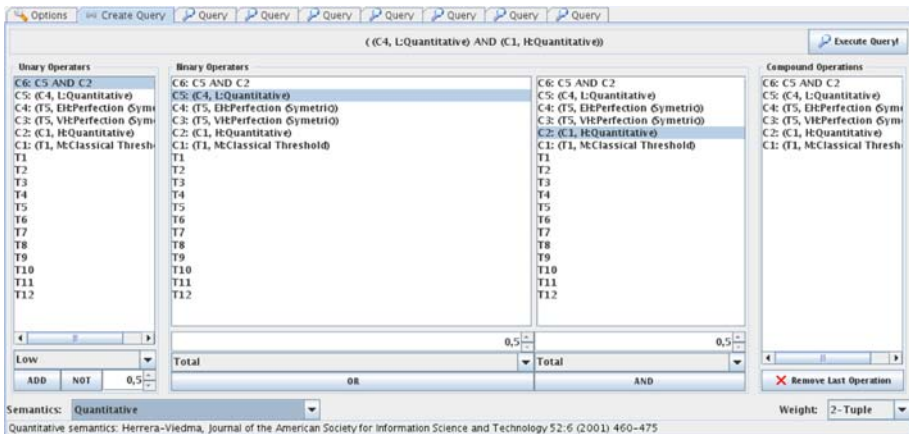

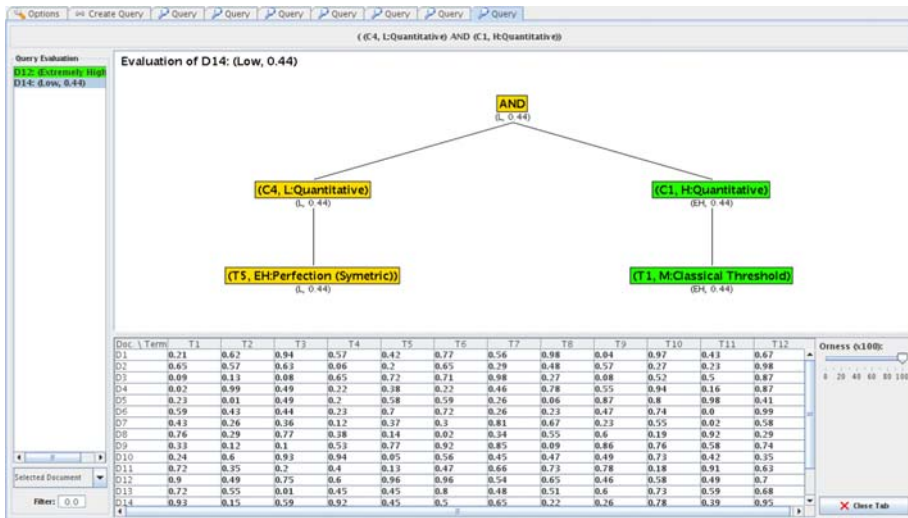**Fig. 6** Defining the weighted query

Fig. 7 Evaluation tree

connective AND. In addition, relevant documents (d_{12} and d_{14}) for the linguistic weighted query q are sorted in decreasing order of $\mathcal{E}$ using linguistic values.

In the following section we present an analysis of the use of our system in the teaching of FIRSs based on weighted queries.

4 Evaluation of the computerized learning system

The computer-supported learning system for FIRSs was tested in the field with students, in order to evaluate its performance and influence on their learning. Although we have some limitations in this study, due to the small number of students (18) participating, its findings provide some interesting insights into student learning and teaching about FIRSs.

Previously, we should point out that in our classes the students are allowed to interact with the system for a maximum of 60 min, 2 days by week during 4 months. This is done under the supervision of the teacher. In addition, students can use the system in their free time from both the computer laboratory or from their house using Internet, (obviously, in those cases they do not have direct supervision from the teacher).

We have used two usual evaluation methods to evaluate the contribution of this computer-supported learning system to the student learning (Alessi and Trollip 1991; Cronje and Fouche 2008; Eteokleous 2008): *a student questionnaire* and *an analysis of exam results*.

4.1 Student questionnaire

We have designed a questionnaire to test the students' opinions after working with our learning system during the course. This questionnaire is composed of four dimensions, two

focused on the exploration of the usability of our learning system and other two focused on its teaching abilities, respectively:

(1)   Interaction
(2)   Interface
(3)   Involvement
(4)   Motivation

   The evaluation criteria in each dimension are adapted for our study from those proposed in (Cronje and Fouche 2008):

(1)   Interaction

   (a)   I felt as if someone was engaged in conversation with me.
   (b)   The feedback was boring.
   (c)   I was given answers, but still do not understand the questions.

(2)   Interface

   (a)   The program is very easy to work with.
   (b)   I did not like the screen layout at all.
   (c)   There are animations in the program that made the contents easy to understand.
   (d)   Sometimes I felt completely lost, the program frustrated me.

(3)   Involvement

   (a)   I prefer the computer based type of lesson to traditional instruction.
   (b)   I was concerned that I might not be able to understand the material.
   (c)   My feeling towards the course material after I had completed the program was favourable.
   (d)   The lessons in the program were dull and difficult to follow.

(4)   Motivation

   (a)   As a result of having studied by this method, I am interested in learning more about the subject matter.
   (b)   I felt quite tense when I worked through the program.
   (c)   I think that what I have learned from the program, should make the normal classroom and laboratory work easier to understand.
   (d)   I think working through the program was a waste of time.
   (e)   The lessons were interesting and really kept me involved.
   (f)   The program challenged me to try my best.

   A nine point checklist format was used to assess the evaluation criteria. A response of 1, 2 or 3 was taken as "disagree", a response of 4, 5 or 6 was taken as "not sure", while a response of 7, 8 or 9 was reported as "agree".
   We have applied this questionnaire in the academic year 2007–2008 in which we worked with 18 students. The questionnaire results are the following:

4.1.1 Interaction

With respect to the dimension "Interaction", pupils' responses indicate that the interaction process with the system was not very good and adequate (see Fig. 8a). About 50% pupils indicated that they disagreed with the statements (a) and (b), and therefore, they found that
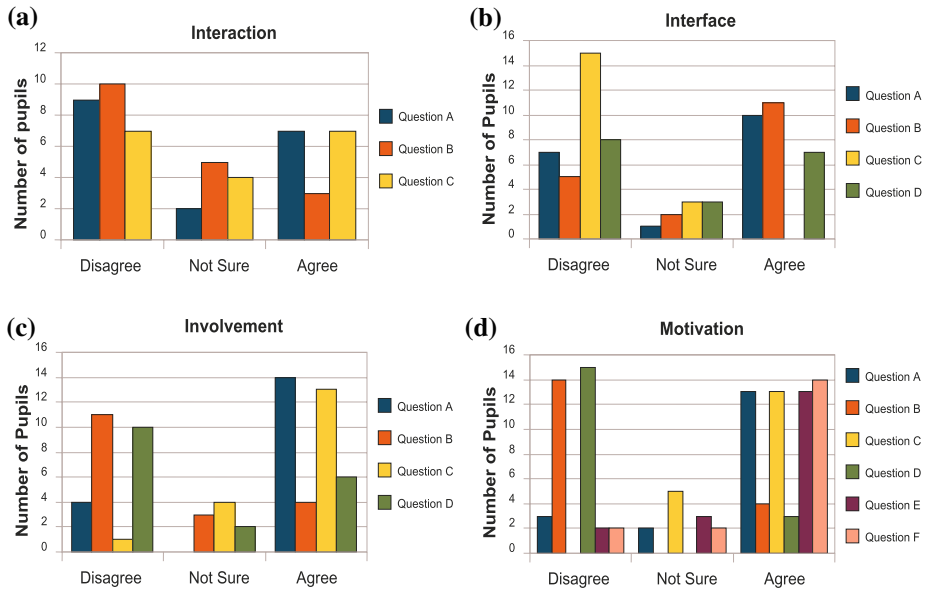
Fig. 8 Results from the questionnaire. **a** Interaction, **b** interface, **c** involvement, **d** motivation

the interaction process was little user-friendly and the feedback was boring. We could improve both aspects of our system using audio/video elements, given that when students were asked why they disagreed the statements, they responded that they "heard nothing" and "would like to see more different types of feedbacks".

On the other hand, in the statement (c) there exists a balance between students' opinions. If we analyze the scores provided by the unsure group of students, {5, 6, 6, 6}, we could understand that the most of students understood the most of answers given to their questions. However, we think that the help module of our learning system should be improved by means of multimedia elements.

4.1.2 Interface

With respect to the dimension "Interface" (see Fig. 8b) we find that the program is easy to use (see criterion (a)) but, however, the majority of pupils did not like the system layout at all and they expected more multimedia components in the learning system. On the other hand, seven students indicated that sometimes they felt completely lost.

Consequently, we think the system's interface could be improved if we re-design the system and incorporate some multimedia instruction elements.

4.1.3 Involvement

In general, after working with our learning system, the most of students responded positively on the involvement dimension (see Fig. 8c): (i) they preferred the computer based type of instruction, (ii) they do not find problems to understand the course material, (iii) they expressed a possitive feeling on the course material, and (iv) they admitted that the

lessons were not difficult to follow neither dull. In a interview with the six students that disagreed with statement (d) they said us that they expected more of an action-type of feedback and that was the reason why they thought that the lessons were dull. Therefore, as above, we think that we should incorporate more multimedia learning elements to facilitate the student's involvement.

4.1.4 Motivation

Similarly, after working with our learning system, the majority of the students answered positively on the motivation dimension. In Fig. 8d we can definitely say that most of the students were positively motivated through the use of our learning system: students showed more interest in the topic of FIRSs, their views on the usefulness of the computer in a classroom changed completely and very few students thought that working through the learning system was a waste of time.

4.2 Analysis of exam results

To get a more objective appreciation on the learning outcome, we have developed two different research studies:

(1) *Does the use of the computer-supported learning system improve the scores between different academic years?*
In this case, we analyze the scores achieved by different groups of pupils across two different academic years in order to study if the computer-supported learning system, as a complement to traditional lectures and exercises, has any positive impact on the learning outcome:

 • Academic year 2005–2006: Pupils that did not use the system, and
 • Academic year 2007–2008: Pupils that used the first stable version of the learning system.[2]

On both groups of pupils, the same methodology with exercises and didactic procedure was carried out along course.

(2) *Does a strong use of computer-supported learning system improve the scores in final exams?*
In this case, we analyze the student scores in academic year 2007–2008 depending on the intensity of the use of the computer-supported learning system.

In both studies, our primary concern for accuracy and to overcome problems with statistical power, is due to the very small sample size ($n = 18$ for academic year 2007–2008 and $n = 16$ for academic year 2005–2006). For such small data sets, it is basically impossible to tell if the data comes from a variable that is normally distributed (Levin and Fox 2006), as with small sample sizes ($n < 20$) tests of normality may be misleading. In this situation, nonparametric tests are an appropriate approach. In addiction, the use of nonparametric methods may be necessary when data has a ranking but no clear numerical interpretation, such as when assessing preferences. In Appendix 2 more details about nonparametric statistical tests are introduced.

---

[2] In academic year 2006–2007 an incomplete and non stable version (β version) was used for the first time. In this sense, scores from this academic year can not be considered as relevant to analyze the learning outcome.

4.2.1 Comparing scores between different academic years 2007–2008 and 2005–2006

The main aim of this research study is the following:

> Was the learning effect on the remembering, understanding and applying level in the pupils from the academic year 2007–2008 higher than in the pupils from the academic year 2005–2006?

To do so, student scores on final exams from both academic years, 2007–2008 and 2005–2006, are directly compared by using the Mann–Whitney's nonparametric statistical test (for more detail see (Mann and Whitney 1947; Sheskin 2003) and Appendix 2).

Then, we analyze the research null hypothesis:

$$H0: Scores_{2007-2008} = Scores_{2005-2006},$$

with $Scores_X$ being the scores on final exams in participants from group X ($X = 2007$–2008 or $X = 2005$–2006). Table 1 shows the scores for participants in groups 2007–2008 and 2005–2006 (0 is the lowest possible score whilst 10 is the maximum one).

The Mann–Whitney's U test (at 95% confidence interval) reflects there is significant difference in scores between group 2007–2008 and group 2005–2006, i.e., $U_{2007-2008} = 202$ ($z = -2,00119$), $U_{2005-2006} = 86$ ($z = 2,00119$), and consequently, the null research hypotheses $H0$ is rejected ($z > 1.96$—see Appendix 2 for more detail). So, given that scores in group 2007–2008 (*median rank* = 18.06) are higher than scores in group 2005–2006 (*median rank* = 12.69) (see Table 1), the alternative hypothesis:

$$Scores_{2007-2008} > Scores_{2005-2006}$$

is supported.

4.2.2 Researching if a strong use of computer-supported learning system improves student scores in the academic year 2007–2008

In order to study if a strong use of the learning system implies higher scores on final exams, a correlation test is applied, i.e, the Spearman's rank correlation test or Spearman's rho -ρ- (see Sheskin 2003 and Appendix 2).

In a correlation test, the correlation value is 1 in the case of an increasing linear relationship, -1 in the case of a decreasing linear relationship, and some value in between in all other cases, indicating the degree of linear dependence between the variables. The closer the correlation is to either -1 or 1, the stronger the correlation between the variables. If the variables are independent then the correlation is 0.

Then, pupils from the academic year 2007–2008 were asked for his/her level of use of the learning system. It was done using a nine point checklist [0: "never used", 1: "sometimes used", …, 8: "only the tool is used for studing/learning (no more materials are used)"]. In Fig. 9, scores and declared use are plotted.

In our study, Spearman's test with $\rho = 0.81$ indicates there is a high correlation (with *p*-value = $4.636e-05$) between scores and declared usage. This represents a statistical statement indicating the presence of an effect or a difference. Such as a positive correlation

Table 1 Student scores

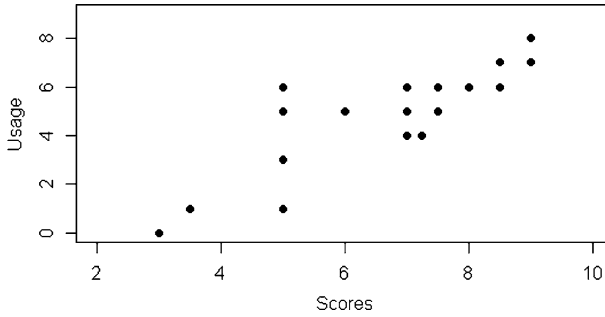|																	
2007–2008	9	8.5	7.5	7	7	3.5	5	6	7.5	5	7	7.25	5	5	3	9	8.5	8
2005–2006	8	8	7.5	7	6.5	3.5	4	6	7	5	3	2	5	5	3	0		

Fig. 9 Plot with scores and declared usage in group 2007–2008

Table 2 Scores and declared use in academic year 2007–2008

Scores	9	8.5	7.5	7	7	3.5	5	6	7.5	5	7	7.25	5	5	3	9	8.5	8
Usage	7	7	6	6	5	1	6	5	5	5	4	4	3	1	0	8	6	6

(and close to 1) reveals that the increase in scores for participants in group 2007–2008 would be a consequence of a strong use of our computer-supported learning system (see values in Table 2).

5 Conclusions

In this contribution a computer-supported learning system to help teachers to teach the use of FIRSs based on weighted user queries has been presented. This system contributes to overcome the teaching problems of IRSs pointed out in (Halttunen and Sormunen 2000), and particularly, the problems to teach weighted query based FIRSs presented at the end of Sect. 2.

We have evaluated its performance on students' learning and our results reveal that the use of this tool enhances students' learning on weighted query based FIRSs, their scores in the final exams, and their involvement and motivation. On the other hand, we have observed that its performance could be improved if we incorporate more multimedia instruction elements in the system activity.

Appendix 1: a fuzzy ordinal linguistic approach

The *ordinal fuzzy linguistic approach* is a fuzzy approximate technique appropriate to deal with qualitative aspects of problems (Herrera and Herrera-Viedma 2000). It models linguistic information by means of ordinal linguistic labels supported by a *linguistic variable*. A linguistic variable is defined by means of a syntactic rule and a semantic rule. In an ordinal fuzzy linguistic approach the syntactic rule is defined by considering a finite and totally ordered label set $\mathcal{S} = \{s_i\}, i \in \{0, \ldots, \mathcal{G}\}$ in the usual sense, i.e., $s_i \geq s_j$ if $i \geq j$, and with odd cardinality (such as 7 or 9 labels), where the mid term represents an assessment of "approximately 0.5", and the rest of the terms being placed symmetrically around it. The semantics of the linguistic term set is established from the ordered structure of the term set by considering that each linguistic term for the pair $(s_i, s_{\mathcal{G}-i})$ is equally informative.

Example A linguistic term set with 9 labels could be $\mathcal{S} = \{s_0 = Null(N),\ s_1 = Extremely\_Low(EL), s_2 = Very\_Low(VL), s_3 = Low(L), s_4 = Medium(M), s_5 = High(H), s_6 = Very\_High(VH), s_7 = Extremely\_High(EH), s_8 = Total(TO)\}$.

In order to manage the linguistic ordinal information we need operators to manage linguistic information, such as:

- *Minimization operator*

$$MIN(s_a, s_b) = s_a\ if\ a <= b.$$

- *Maximization operator*

$$MAX(s_a, s_b) = s_a\ if\ a >= b.$$

- *Negation operator*

$$NEG(s_i) = s_j\ |\ j = \mathcal{G} - i.$$

- *Aggregation operators* we can use aggregation operators based on symbolic computation, as for example the Linguistic OWA operator (Herrera et al. 1996) or the Linguistic Weighted Averaging operators (Herrera and Herrera-Viedma 1997).

Appendix 2: introduction to inferential statistical tests

This section briefly introduces the necessary issues to understand the statistical terms used in this paper, i.e, Mann–Whitney's U test and Spearman's Rank Correlation Coefficient.

Mann–Whitney U test

This test was developed independently by Wilcoxon in 1945 and Mann and Whitney in 1947 (Sprent 1993). Therefore, it is known by two different names: the Wilcoxon rank-sum test and the Mann–Whitney U test. The Wilcoxon rank-sum test only explores equal sample sizes. The test proposed by Mann and Whitney, which is an extension of the Wilcoxon rank-sum test, resolved this problem.

The Mann–Whitney U test assesses the null hypothesis, which states that the two samples come from identical distributions. Also, the Mann–Whitney U test requires the samples to have continuous measurements (Mann and Whitney 1947). For two samples X and Y, the Mann–Whitney U test first ranks the two samples. To do so, the samples are sorted into one large sample while maintaining the information about which sample each value came from. Each value is then ranked with a number between 1 and $|X| + |Y|$. If there are two or more identical values, the average of the ranks of these values is assigned as their identical ranks. For example, if k identical values $v_1,\ldots,v_k$ have ranks $r_1,\ldots,r_k$, they all get the same rank: $\frac{\sum_{i=1}^{k} r_i}{k}$.

It is clear that ranking the values removes the effects of the outliers present in certain tests such as parametric tests. A value large enough to be considered an outlier will have a larger rank than the previous non-outlier value in the sorted list. However, this rank is only

larger by a constant $c < k$. Once all the values of both samples are ranked, the Mann–Whitney U test calculates the sum of the ranks for all the values in each sample. Therefore, for a sample X, the sum of the ranks is:

$$R_X = \sum_{r_i \in X} r_i.$$

R_Y is calculated the same way.

Using the sum of the ranks for each sample, the U statistic is calculated as follows, where $n_X = |X|$:

$$U_X = R_X - \frac{n_X(n_X + 1)}{2}.$$

U_Y is calculated the same way. In addition to U_X and U_Y, the mean and standard deviation of U must also be found using the following equations, where m_U is the mean of U, σ_U is the standard deviation of U, and $n_Y = |Y|$:

$$m_U = \frac{n_X n_Y}{2},$$

$$\sigma_U = \sqrt{\frac{n_X n_Y (n_X + n_Y + 1)}{12}}.$$

Finally, the following normal approximation can be used to calculate the standard normal deviate, z, also known as the standard score:

$$z = \frac{U - m_U}{\sigma_U}.$$

Since there are two samples X and Y, there will be two standard scores, z_X and z_Y. These values are always equal in magnitude and opposite in sign. Therefore, they are simply identified by z, where $z = |z_X| = |z_Y|$. In order to assess the null hypothesis, a confidence interval is used. At 95% confidence interval, $z <= 1.96$ must hold to accept the null hypothesis.

Given two methods m_X and m_Y, the acceptance or rejection of the null hypothesis can only identify these methods' performance as being identical or not respectively. In the case of non-identical methods, the test does not help in identifying which method performed better. Nonetheless, in comparison of different methods it is more interesting to find the ones that performed best. Therefore, in case of two non-identical samples, which happens when the null hypothesis is rejected, another non-parametric test can augment the Mann–Whitney U test. The simplest and most useful non-parametric test is to compare the median ranks of the two samples after the Mann–Whitney U test considered them non-identical. In the cases where higher ratings imply better quality, if sample X has a higher median rank than sample Y, then method m_X is considered better than method m_Y.

Spearman's rank correlation coefficient

In statistics, Spearman's rank correlation coefficient or Spearman's rho (Spearman 1904), named after Charles Spearman and often denoted by the Greek letter ρ, is a non-parametric measure of correlation that is, it assesses how well an arbitrary monotonic function could describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables.

In principle, ρ is simply a special case of the Pearson product-moment coefficient in which two sets of data X_i and Y_i are converted to rankings x_i and y_i before calculating the coefficient. In practice, however, a simpler procedure is normally used to calculate ρ. The raw scores are converted to ranks, and the differences d_i between the ranks of each observation on the two variables are calculated (Myers and Well 2006).

If there are no tied ranks, i.e.

$$\neg \exists_{i,j}(i \neq j \wedge (X_i = X_j \vee Y_i = Y_j))$$

then ρ is given by: $\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$ where $d_i = x_i - y_i$ is the difference between the ranks of corresponding values X_i and Y_i, and n is the number of values in each data set (same for both sets).

If tied ranks exist, classic Pearson's correlation coefficient between ranks has to be used instead of this formula:

$$\rho = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum x_i^2) - (\sum x_i)^2}\sqrt{n(\sum y_i^2) - (\sum y_i)^2}}.$$

One has to assign the same rank to each of the equal values. It is an average of their positions in the ascending order of the values.

References

Alessi, S. M., & Trollip, S. R. (1991). *Computer-based instruction: Methods and development* (2nd ed.). Englewood Cliffs, NJ: Prentince-Hall.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison-Wesley.

Bonissone, P. P. (1997). Soft computing: The convergence of emerging reasoning technologies. *Soft Computing, 1*(1), 6–18.

Bordogna, G., & Pasi, G. (1993). A fuzzy linguistic approach generalizing boolean information retrieval: A model and its evaluation. *Journal of the American Society for Information Science, 44*(2), 70–82.

Buell, D., & Kraft, D. H. (1981). Threshold values and boolean retrieval systems. *Information Processing & Management, 17*, 127–136.

Caruso, E. (1981). Computer aids to learning online retrieval. *Annual Review of Information Science and Technology, 17*, 317–336.

Chau, M., Huang, Z., & Chen, H. (2003). Teaching key topics in computer science and information systems through a Web search engine project. *ACM Journal of Educational Resources in Computing, 3*(3), 1–14.

Chiclana, F., Herrera-Viedma, E., Herrera, F., & Alonso, S. (2004). Induced ordered weighted geometric operators and their use in the aggregation of multiplicative preference relations. *International Journal of Intelligent Systems, 19*, 233–255.

Chiclana, F., Herrera-Viedma, E., Herrera, F., & Alonso, S. (2007). Some induced ordered weighted averaging operators and their use for solving group decision-making problems based on fuzzy preference relations. *European Journal of Operational Research, 182*(1), 383–399.

Crestani, F., & Pasi, G. (2000). *Soft Computing in Information Retrieval: Techniques and applications. Studies in Fuzziness and Soft Computing Series* (Vol. 50). Physica-Verlag.

Cronje, J. C., & Fouche, J. (2008). Alternatives in evaluating multimedia in secondary school science teaching. *Computers & Education, 51*, 659–683.

Eteokleous, N. (2008). Evaluating computer technology integration in a centralized school system. *Computers & Education, 51*, 669–686.

Griffith, J. C., & Norton, N. P. (1981). A computer assisted instruction program for end users on an automated information retrieval systems. In *Proceedings of the second national online meeting*, pp. 239–248, New York.

Halttunen, K., & Sormunen, E. (2000). Learning information retrieval through an educational game. Is gaming sufficient for learning? *Education for Information, 18*(4), 289–311.

Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instructions, 14*, 343–351.

Herrera, F., & Herrera-Viedma, E. (1997). Aggregation operators for linguistic weighted information. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 27*, 646–656.

Herrera, F., & Herrera-Viedma, E. (2000). Linguistic decision analisys: Steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems, 115*, 67–82.

Herrera, F., Herrera-Viedma, E., & Verdegay, J. L. (1996). Direct approach processes in group decision making using linguistic OWA operators. *Fuzzy Sets and Systems, 79*, 175–190.

Herrera-Viedma, E. (2001a). An information retrieval system with ordinal linguistic weighted queries based on two weighting elements. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 9*, 77–88.

Herrera-Viedma, E. (2001b). Modelling the retrieval process for an information retrieval system using an ordinal fuzzy linguistic approach. *Journal of the American Society for Information Science and Technology, 52*(6), 460–475.

Herrera-Viedma, E., & López-Herrera, A. G. (2007). A model of information retrieval system with unbalanced fuzzy linguistic information. *International Journal of Intelligent Systems, 22*(11), 1197–1214.

Herrera-Viedma, E., Cordón, O., Luque, M., López, A. G., & Muñoz, A. M. (2003). A model of fuzzy linguistic IRS based on multi-granular linguistic information. *International Journal of Approximate Reasoning, 34*, 221–239.

Herrera Viedma, E., López Herrera, A. G., & Porcel, C. (2005). Tuning the matching function for a threshold weighting semantics in a linguistic information retrieval system. *International Journal of Intelligent Systems, 20*, 921–937.

Herrera-Viedma, E., Pasi, G., & Crestani, F. (2006). *Soft computing in Web Information Retrieval: Models and applications. Studies in fuzziness and soft computing series* (Vol. 197). Physica-Verlag.

Herrera-Viedma, E., López-Herrera, A. G., Luque, M., & Porcel, C. (2007). A fuzzy linguistic IRS model based on a 2-tuple fuzzy linguistic approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15*(2), 225–250.

Hull, D. (1996). Stemming algorithms: A case study for detailed evaluation. *Journal of the American Society for Information Science, 52*(6), 70–84.

Kraft, D. H., Bordogna, G., & Pasi, G. (1994). An extended fuzzy linguistic approach to generalize boolean information retrieval. *Information Sciences, 2*, 119–134.

Levin, J., & Fox, J. (2006). *Elementary statistics in social research*. Boston: Allyn & Bacon.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics, 18*(1), 50–60.

Markey, K., & Atherton, P. (1978). *ONTAP online training and practice manual for ERIC data base searchers*. Syracuse, New York: ERIC Clearinghouse on Information Sources, Syracuse University.

Myers, J. L., & Well, A. D. (2006). *Research design and statistical analysis* (2nd ed.). Lawrence Erlbaum.

Nikravesh, M., Loia, V., & Azvine, B. (2002). Special issue on fuzzy logic and the internet (flint): Internet, world wide Web and search engines. *Soft Computing, 6*(5), 287–299.

Salton, G., & McGill, M. J. (1983). *An introduction to Modern Information Retrieval*. McGraw-Hill.

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*. CRC Press.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal or Psychology, 15*, 72–101.

Sprent, P. (1993). *Applied nonparametric statistical methods*. London: Chapman & Hall.

Stratford, S. (1997). A review of computer-based model research in precollege science classrooms. *Journal of Computers in Mathematics and Science Teaching, 16*(1), 3–23.

Waller, W. G., & Kraft, D. H. (1979). A mathematical model of a weighted boolean retrieval system. *Information Processing & Management, 15*, 235–245.

Yager, R. R. (1987). A note on weighted queries in information retrieval systems. *Journal of the American Society of Information Sciences, 38*, 23–24.

Yager, R.R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man, and Cybernetics, 18*, 83–190.

Yager, R. R., & Filev, D. P. (1999). Induced ordered weighted averaging operators. *IEEE Transaction on Systems, Man and Cybernetics, 29*, 141–150.

Yaman, M., Nerdel, C., & Bayrhuber, H. (2008). The effects of instructional support and learner interests when learning using computer simulations. *Computers & Education, 51*, 1784–1794.