Fuzzy Group Models for Adaptation in Cooperative Information Retrieval Contexts

Miguel-Ángel Sicilia and Elena García

Computer Science Department. Polytechnic School.
University of Alcalá. Ctra. Barcelona km. 33.6
28871 – Alcalá de Henares, Madrid (Spain)
{msicilia,elena.garciab}@uah.es

Abstract. Cooperation in information retrieval contexts can be used to share query results inside groups of individuals with common objectives, provided that all of them are aware of each other. The strength of the social relationships between group members is in most cases a matter of comparative degree, and thus relationships can be modelled through fuzzy conceptual associations. These associations can then be used to implement personalized features, aimed at improving the interaction of the user with the query tool. In this paper, an approach to modelling imprecise relationships between users in the context of information retrieval is described, along with a concrete case study implemented as a wrapper of a conventional search engine, using a fuzzy database to store the model of the group members.

1   Introduction

Research on *interactive information retrieval* (IIR) [16] is concerned with the study of human interaction issues in their process of using information retrieval (IR) systems. IIR can be considered a complementary view of the classical system-oriented view of IR (see, for example, [8]), and has given rise to specific models like the interactive feedback and search model proposed by Spink [23]. From among the diverse complex interactions that are pursued in IR contexts, cooperation[1] between users in a group have resulted in specific algorithms that exploit the history of queries (see, for example, [9]) in the process of resolving new ones. But, in addition to improved retrieval algorithms, the social context of the search activity can be exploited to improve other aspects of user's every-day's search tasks when the group is organized around common objectives. For

---

[1] We use here the term cooperation: "to act or work with another or others" as a more general term than collaboration "to work jointly with others or together ", since the latter entails a common concrete desired outcome.

example, in the context of a research group, users can share some of their conference bookmarks that are supposedly of interest to other people in the group. In consequence, user interfaces for IR systems that enable cooperation [17] can take advantage of some kind of *group model* — i.e. a model of the interrelation between group members — to implement adaptive features [5].

Collaborative filtering (CF) systems [11] exploit similarities between users in browsing or search processes, but they do that in a transparent manner for the user, so that people receiving recommendations are unaware of the fact that neighborhoods of users that somewhat agree with them have been computed by the CF algorithm. Thus, a system-oriented view of groups is used, that does not take into account the potential of explicit, user-initiated interactions between users that are socially aware of the others, which are known to improve cooperation, as studied in systems like [13].

In this paper, a concrete cooperative IR model is described, which provides a number of features for the explicit cooperation between users in a group with related interests (i.e. that provides functionalities for cooperative search *tactics* [2]). Two key aspects of such a cooperative setting are the modelling of the *closeness* between users and of the *perceived relevance* of certain types of communication events like user-to-user recommendations. Both kinds of social relationships between group members are inherently imprecise and uncertain, since they evolve with time and they should be measured by some form of analysis of the recorded interaction history. We describe here how these relationships can be modelled as fuzzy relationships at the conceptual level, and also how they can be stored in fuzzy relational database structures in the final implementation. In addition, relationships can be used to define fuzzy categories of users with regards to their social interaction characteristics, like the concept of "pertinent" users. Although the model only deals with fuzziness at the group model level, it could be potentially combined with existing fuzzy techniques that enhance the query resolution process — see, for example [12].

The rest of this paper is structured as follows. In Section 2, the general cooperative information retrieval setting is described. Section 3 provides the conceptual description of a concrete case study that uses fuzziness to model social relationships, along with a brief description of the implementation. Finally, conclusions and future research directions are provided in Section 4.

2   Modelling Fuzzy Social Relations in Group IR Tasks

Cooperative IR contexts are used by one or several groups of users that can be denoted by a set $G = \{g_1, \ldots, g_n\}$. Each of the groups is made up of a collection of users from a set U that can be denoted as $users(g_i)$ or $g_i = \{u_1^i, \ldots, u_m^i\}$. Communications can be initiated by users inside the same group. *Awareness* is taken for granted in this model, so that it's assumed that the group constituents know each other *a priori* or appropriate awareness tools are provided.

The information retrieval system $\mathcal{S}$ continuously resolves queries matching query strings against the contents of its resource base, which essentially is a set

of contents C (in different, possibly multimedia, formats). The set representing the history of queries of the group g_i is denoted by s_{g_i}, and the subset of queries for a given user $x \in g_i$ is denoted by $s_{g_i}^x$.

Essentially, two basic policies can be used to improve the interaction of each user from information provided by others users of the group. *System-intentional* interactions take information from the group constituents to obtain patterns or similarity measures. Research on *collaborative filtering* is mainly focused on this kind of interaction — even when it target groups like in [15] —, in which the system elaborates recommendations by exploring relationships between items and between users [18]. *Peer-intentional* interactions are of a very different nature. In this case, the users consciously direct suggestions or indications — that they believe to be interesting or significant — to some of their colleagues. Of course, both kinds of policies can be combined.

The system provides a number of peer-intentional interactions in the form of asynchronous messages with signatures, i.e., $\mathbf{m}(parameters)$. A typical example of such a message is that of recommending a search result to a peer or a set of them, which could be modelled as $\mathbf{recommend}(u : 2^U; \ c : C)$. This concrete type of message is central to the case study described below.

The general form of a social relationship inside a group can be denoted by a fuzzy relation with the form $R : g_i \times g_i \rightarrow [0, 1]$, so that $R(u_a^i, u_b^i)$ represents the strength of the relationship between users a and a inside group i. The model of each social relationship $R \in \mathcal{R}$ that we are concerned in this work is thus constructed from the history of message instances by some form of computation algorithm denoted by $\mathcal{C}_R$. In addition, fuzzy categories of users can be defined as fuzzy subsets of the set of users in the social group $g_i \in G$ being considered. Both forms of fuzziness are generalizations of their crisp counterparts that are found in adaptive hypermedia modelling frameworks [21].

3 Case Study Description

The generic modelling approach described in Section 2 has been used for the design of a research group-oriented cooperative IR interface called `deiSearch`, which follows a peer-intentional policy of interaction. The application is essentially a wrapper using the `Google` programming interfaces[2] to resolve queries and obtain results, while recording the search history of each group member to implement adaptive features based on simple cooperative behaviors.

3.1 Conceptual Model

The conceptual modelling elements for adaptive hypermedia applications described in [21] can be used to specify the model of `deiSearch` in term of the major architectural components usually found in adaptive hypermedia, known as user, domain and adaptation model.

---

[2] [http://www.google.com/apis/]

Figure 1 shows the essential elements of the model and its relationships in the form of a UML [14] static diagram.
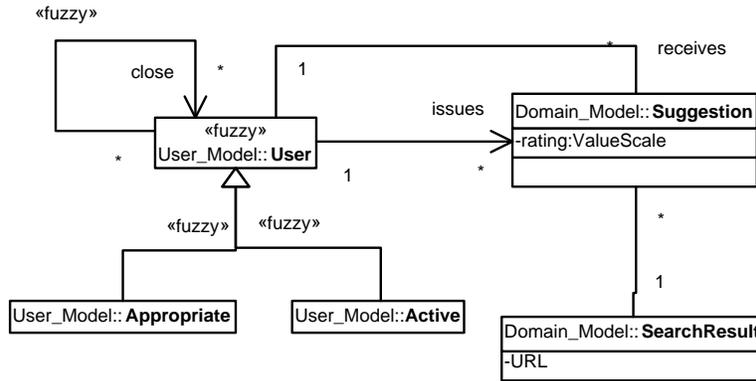

Fig. 1. Core Elements of the Case Study Model, depicted as a UML diagram

The model in Figure 1 is structured around packages corresponding to common high-level architectural components of adaptive hypermedia systems [25], namely, the user model (in our case, a group model) and the domain model (that is, the terms that are specific of the application's domain). Essentially, users interact with the IR system by issuing queries that return sets of search results, and they are allowed to share some of these results with other users, as represented by the `Suggestion` class in Figure 1. These suggestions are subject of rating by its receivers, using a linguistic label set (represented by the `ValueScale` data type) for that purpose. The following modelling notions about users have been considered for the application:

- The concept of user, embodied in the `User` class is considered as a fuzzy class or set (denoted by the `<<fuzzy>>` UML stereotype[3]). The grade of membership to this class is directly computed from the frequency of use of the application.
- Closeness between users is modelled as a fuzzy relation, represented as a `<<fuzzy>>` association (see [19]). This relationship is constructed from the history of recommendations, that serves as an estimator due to the fact that deeper knowledge of each other is assumed to be correlated to the quantity of interactions.
- A fuzzy subset `Appropriate` of the users is defined. This relationship models a form of trust — in the sense given in [1] — that is essentially a (inter-)subjective belief about the usefulness of the recommendations of a given individual. This kind of subsets are modelled by a `<<fuzzy>>` generaliza-

---

[3] Stereotypes are a way of defining virtual, specialized meta-classes in UML

tion/specialization relationship, which defines a fuzzy subset between two classifiers, so that if B is a subclass of A, $\forall x\ \mu_B(x) \leq \mu_A(x)$.

– A fuzzy subset `Active` of the users is also defined (again as a fuzzy specialization). This subset is a characterization of the level of communicative activity (in our case, the frequency of recommendations issued by him/her). It should be noted that a user may be very active, but also very unappropriate, that is, it may issue a large number of messages that are perceived grossly irrelevant by his/her peers.

The degree of belonging to the `User` class is obtained in a straightforward manner by the computation described by the expression (1), where a group $g \in G$ is assumed, so that s_g is the set of queries of the group and s_g^x is the subset of queries of user x. Variable t is a temporal adjustment factor to model the fact that in initial stages of the algorithm, the percentage of searches initiated by a user affects only slightly its belonging to the user class. It starts as a large number that can be adjusted to each concrete setting (e.g. 10^q with $q = 4$) so that q decreases till reaching zero and its effect disappears after a reasonable and preestablished time has elapsed.

$$\mu_{user}(x) = min(1,\ [1 - \frac{\sum_{u \in g-\{x\}} |s_g^u|}{|s_g|} \cdot t] \cdot k) \tag{1}$$

The constant k should be somewhat dependant of the size of the group, since in bigger groups, the same percentage of participation must be considered as a larger degree of evidence about system's usage. Concretely, we have used the $\frac{|g|}{2}$ value, which provides a good heuristic for groups that are below 30 participants. For example, in a group of twenty participants, after t has reached its top, a member with a five percent of the total number of searches is considered as having a 0.5 degree of 'userness'. But with the same k definition, in a group of a hundred participants, an individual with that five percent of searches would be considered as fully belonging to the (fuzzy) set of users. Note that the expression (1) only assigns partial membership grades to users that have non-significant volumes of activity. Table 1 provides a sample of usage data. Concretely, it provides the number of queries issued by a group of ten researchers in intervals of ten days during two months.

Figure 2 depicts the corresponding evolution of the degree of *userness* of six of the users in Table 1 with an initial $q = 2$. It should be noted that all of them are provided initially with a full belonging to that set, which models the initial uncertainty of the system derived from the small amount of evidence available.

The subgroup of *active* users is computed by the number of recommendation messages issued by each individual, in a similar way to the degree of 'userness', but under the constraint that $\mu_{user}(x) \geq \mu_{active}(x)$.

The degree of appropriateness of a user is computed from his/her history of recommendations by the expression (2). The computation algorithm $\mathcal{A}_{appropriate}$ operates in two phases:

– First, each of the peer assessments v_s^u, where $u \in g - \{x\}$, of the suggestions of the user x under consideration are aggregated by an aggregation operator

Table 1. Example query data for a group of ten users.

user	10	20	30	40	50	60
a	10	54	38	36	15	20
b	60	84	92	115	93	88
c	32	100	54	45	90	56
d	30	32	36	38	43	34
e	45	43	21	34	54	56
f	67	56	46	58	30	34
g	34	29	12	2	12	82
h	10	12	14	16	23	25
i	12	34	40	20	16	21
j	2	11	81	34	9	12
totals	302	455	434	398	385	428

$\mathcal{M}_1$ (satisfying the general properties of this kind of elements described in [4]), as showed in (3).

– After that, a second aggregation operator $\mathcal{M}_2$ (2) is used to combine the assessments of all the suggestions issued by the user. In this case, the degree of *activeness* described before can be used to model the fact that not only the adequacy but the volume of suggestions are considered in our concept of 'appropriateness', since it's intended to reflect some form of utility of the individual with regards to the objectives of the entire group.

$$\mu_{appropriate}(x) = \mathcal{M}_2(v(s_1), \ldots, v(s_p)) \cdot \mu_{active}(x) \tag{2}$$

$$\forall s \in Suggestions \quad v(s) = \mathcal{M}_1(v_s^{u_1}, \ldots, v_s^{u_r}) \tag{3}$$

A natural choice for the operators $\mathcal{M}_1$ and $\mathcal{M}_2$ would be that of an OWA operator modelling majority [26], but other design consideration would also be considered in concrete settings, like, for example, different weights assigned to individuals depending on their *reputation*[1]. Empirical assessments can also be used for the determination of the most appropriate quantifier design (see, for example, [24]). For example, let's suppose that user f (having a degree of activeness of 0.32 at t=60) issued eight suggestions that have resulted in a set of aggregated values $\{5, 4, 5, 3, 1, 5, 4, 2\}$ (using $\mathcal{M}_1$). If we use a OWA operator with weighting vector $w = (0.3, 0.2, 0.2, 0.1, 0.05, 0.05, 0.05, 0.05)$ (which has a degree of *orness* of about 0.143) as $\mathcal{M}_2$, the result of the aggregation will yield a value of 4.4, significantly higher than that of a simple weighted mean (this *orness*-related plasticity of OWA operators allow for iterative adjustments of the aggregation procedure). Then, the final appropriateness level of the user would be $\mu_{appropriateness} = 0.88 \cdot 0.32$, with the result of the OWA operator normalized into the $[0, 1]$ interval.

Finally, the (symmetric) degree of closeness is computed by the expression given in (4), where the amount of messages received is denoted in each direction
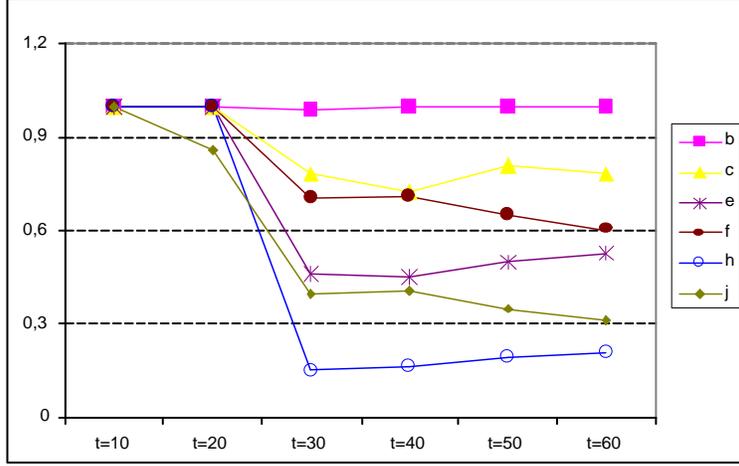
Fig. 2. Evolution of the degree of userness for six of the users

are denoted by $m_{i \rightarrow j}$ and $m_{j \rightarrow i}$, and k has the same purpose as that defined in (1).

$$\forall i \neq j \ \ closeness(i,j) = min(1, min(\frac{m_{j \rightarrow i}}{m_{i \rightarrow j}}, \frac{m_{i \rightarrow j}}{m_{j \rightarrow i}}) \cdot \frac{m_{j \rightarrow i} + m_{i \rightarrow j}}{\sum_{x,y \in g \times g} m_{x \rightarrow y}} \cdot k) \ \ (4)$$

The minimum of the relation of the messages between i and j in both directions is used to filter unidirectional relations, which are not good estimators for closeness. For example, if i issues a hundred messages to j but receive only one from i, the minimum will yield a low value, while the maximum value will be obtained with the amount of messages is equal in both directions, which indicates a constant cooperation. The second part of the formula in (4) is used as an heuristic to increase closeness for the most active pairs of users, so that a pair of users with a total amount of messages interchanged. For example, if we have a total number of 178 messages interchanged in the group, and given that $m_{h \rightarrow c} = 26$ and $m_{c \rightarrow h} = 18$, we have a degree of closeness $closeness(h,c) = min(1, 0.692307692 \cdot \frac{26+18}{178} \cdot 5) = 0.856$.

It should be noted that the computation algorithms are not intended to be general models for cooperative IR, but concrete realizations that embody specific, designer-dependent characteristics, so that other formulations are possible.

Based on the models for the social aspects just described, the computation algorithms are implemented as processes triggered in a periodical basis that updates a relational database. The fJDBC framework described in [20] was used for the logical database design derived from the conceptual model described above. This model provides direct support for the representation of fuzzy generalization

relationships, fuzzy classes and fuzzy associations, and enables direct querying from JDBC[4]-like code.

3.2 Adaptive Behaviors

The overall interface of `deiSearch` is fairly similar to the basic `Google` interface. In Figure 3, a screenshot of the results of a query is shown. It should be noted that a link under each result is provided to issue a recommendation message.
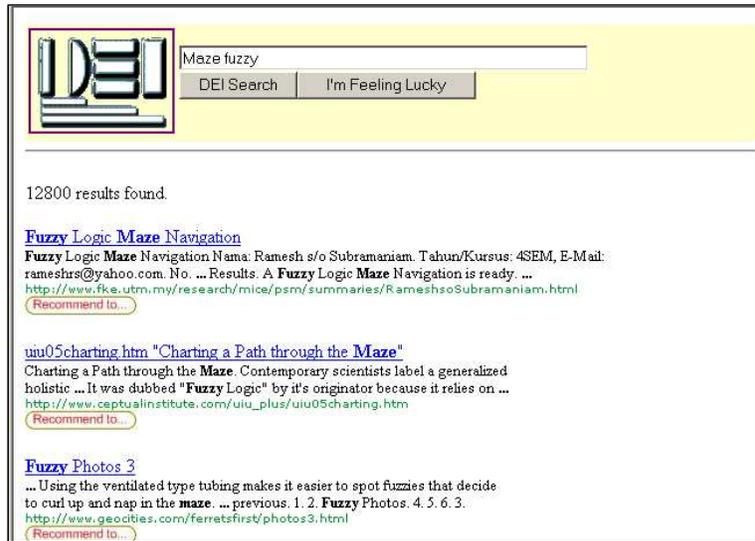

Fig. 3. Sample `deiSearch` search results interface

The following *adaptive technologies* [3] that use the above-described fuzzy categories and relationships have been implemented:

- When elaborating a recommendation, the user is presented with a list of the members of the group. The closeness relationship is used to implement adaptive sorting, as showed in Figure 4 (note that the grades are showed only for illustration and they're normally hidden to the user). This entails that the user would more likely find the intended receivers without a need to scroll or read the entire list.
- When a user is receiving new recommendations (by traversing a link in the principal interface), an adaptive sorting process is used again, but this time using the overall appropriateness degree of the issues, so that the most 'trustworthy' individuals are considered first.

---

[4] The `Java` API for database access: [http://java.sun.com/products/jdbc/]

– In the case that a user receives a recommendation that comes from a highly relevant user that, simultaneously, has a high degree of closeness to him, a *direct navigation* adaptive feature is used, so that a link to the recommendation is put directly in the principal interface, so that it would very likely be read immediately by the receiver.

Figure 4 shows an example of how adaptive sorting based on the degree of closeness between the current user and the rest of the users.

Recommending item:

Fuzzy Logic Maze Navigation

	User	Compatibility
☐	elena.garciab	0.6
☐	aedo	0.4
☐	pdp	0.4
☐	jmdodero	0.2

Fig. 4. Sample adaptive sorting in `deiSearch`

The adaptation is performed using `fJDBC` queries which are syntactically `SQL` sentences that are translated internally to compute the degrees of membership to fuzzy notions. For example, a query to retrieve the (fuzzy) subset of active users is as simple as `SELECT * FROM ACTIVE_USERS`, and the definition of the `ACTIVE_USERS` table is done as described in [20]:

```
<fuzzy:fuzzy-relation name="ACTIVE_USERS" type="subclass">
      <fuzzy:relation  name="users"/>
      <fuzzy:oidField  name="userid"/>
      <fuzzy:membField name="Muser"/>
  </fuzzy:fuzzy-relation>
```

The `Tomcat`[5] Java-enabled Web server was used to implement the just described case study.

Initial evaluations of the prototype in a group with ten participants for two months pointed out that a significant increment (about 15%) in query results considered as relevant for the task (i.e. searching for information about a specific research theme oriented toward writing a project proposal). Nonetheless, further methodological and experimental work is required in evaluating the effects of the cooperation-oriented and adaptive features.

---

[5] http://jakarta.apache.org/tomcat/

4 Conclusions and Future Work

Social relationships in cooperative IR settings can be modelled by fuzzy associations and fuzzy categories constructed from the history of communications between users. A concrete case study has been described that models the notion of closeness and of relevance of recommendations. These kind of relationships enable the development of adaptive features that enhance explicit cooperation tactics between users. In addition, an underlying fuzzy relational schema has been used for the implementation.

Future work should carry out additional inquiries about which kind of relationship are the most appropriate for specific work contexts, in order to customize the overall framework described here to concrete objectives. In addition, the use of fuzzy relationships in user collaboration mechanisms for specific tasks [6] will be considered. These improvements should take into account cross-cultural studies like [7], and also long-term realistic revision mechanisms [22] for the group model.

The interpretation given to imprecision and uncertainty of social relationships in the described case study has been chosen by its simplicity and according to design heuristics, but it may be approached from more elaborated viewpoints with regards to uncertainty, e.g. from belief-theory or other related frameworks [10].

References

1. Abdul-Rahman, A. and Hailes, S.: Relying on Trust to Find Reliable Information. In: Proceedings of the 1999 International Symposium on Database, Web and Cooperative Systems (DWACOS'99), Baden-Baden, Germany
2. Bates, M.J.: Where Should the Person Stop and the Information Search Interface Start?. Information Processing & Management 26 (1990): 575–591
3. Brusilovsky, P.: Adaptive hypermedia. User Modeling and User Adapted Interaction 11(1/2) (2001): 87–110
4. Calvo, T. et al.: Aggregation Operators: Basic Concepts, Issues and Properties. In: Calvo, T., Mayor, G. and Mesiar, R. (eds.). Aggregation Operators: New Trends and Applications. Studies in Fuzziness and Soft Computing, 97 (2002):3–106
5. Davis, A. and Bueno, D.: User modeling and cooperative information retrieval in information retrieval systems. International Journal of Knowledge Organization 26(1) (1999):30–45
6. Dodero, J.M., Aedo, I. and Díaz, P.: Participative knowledge production of learning objects for e-books. Electronic library, 20(4), (2002):296–305.
7. Duncker, E: Cross-cultural usability of the library metaphor. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (2002): 223–230
8. Grossman, D.A. and Frieder, O.: Information Retrieval: Algorithms and Heuristics. Kluwer (1998)
9. Hust, A., Klink, S., Junker, M. and Dengel, A.: Towards Collaborative Information Retrieval: Three Approaches. In: Franke, Nakhaeizade, Renz (eds.): Text Mining - Theoretical Aspects and Applications, Springer-Verlag (2002)
10. Klir, G.J. and Wiermann, M.J.: Uncertainty-Based Information. Elements of Generalized Information Theory. Physica-Verlag, Heidelberg (1998)

11. Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J.: GroupLens: Applying Collaborative Filtering to Usenet News. Communications of the ACM 40(3) (1997): 77–87

12. Kraft, D.H., Bordogna, G., and Pasi, G.: Fuzzy Set Techniques in Information Retrieval. In: Bezdek, J.C., Didier, D. and Prade, H. (eds.), Fuzzy Sets in Approximate Reasoning and Information Systems, vol. 3, The Handbook of Fuzzy Sets Series, Norwell, MA: Kluwer Academic Publishers (1999):469–500

13. Liechti, O., Sifer, M.J. and Ichikawa, T.: A Non-obtrusive User Interface for Enabling Social Awareness on the World Wide Web. Personal Technologies 3(3) (1999): 22–32

14. Object Management Group (OMG): The Unified Modelling Language Specification, version 1.4, September 2001

15. O'Connor, M., Cosley, D., Konstan, J. A. and Riedl, J.: PolyLens: A Recommender System for Groups of Users. In: Proceedings of the seventh European Conference on Computer Supported Cooperative Work (2001):199–218

16. Robins, D.: Interactive Information Retrieval: Context and Basic Notions. Informing Science Journal 3(2) (2000)

17. Romano, N., Roussinov, D., Nunamaker, J.F. and Chen, H.: Collaborative information retrieval environment: integration of information retrieval with group support systems. In: Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS-32) (1999)

18. Sarwar, B. M., Karypis, G., Konstan, J.A. and Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International World Wide Web Conference (2001):285–295

19. Sicilia, M.A., Gutiérrez, J.A. and García, E.: Designing Fuzzy Relations in Orthogonal Persistence Object-Oriented Database Engines. In: Proceedings of the 8th Ibero-American Conference on AI IBERAMIA'02, Springer Lecture Notes in Computer Science 2527 (2002):243–253

20. Sicilia, M.A., García, E., Díaz, P. and Aedo, I.: Extending Relational Data Access Programming Libraries for Fuzziness: The fJDBC Framework. In: Proceedings of the 5th International Conference on Flexible Query Answering Systems. Lecture Notes in Computer Science 2522, Springer (2002):314–328

21. Sicilia, M.A., García, E., Díaz, P. and Aedo, I.: Fuzziness in Adaptive Hypermedia Models. In: Proceedings of the North American Fuzzy Information Processing Society (2002):268–273.

22. Sicilia, M.A. Observing Web Users: Conjecturing and Refutation on Partial Evidence. In: Proceedings of the North American Fuzzy Information Processing Society (2003):530–535.

23. Spink, A. and Saracevic, T.: Human computer interaction in information retrieval: Nature and manifestations of feedback. Interacting with Computers: The International Journal of Human Computer Interaction, 10(3) (1998): 249–267

24. Torra, V.: Empirical analysis to determine Weighted OWA orness. In: Proceedings of the 5th Intl. Conference on Information Fusion (Fusion 2001), Montreal, Quebec, Canada (2001)

25. Wu, H., De Kort, E. and De Bra, P.: Design issues for general-purpose adaptive hypermedia systems. In: Proceedings of the ACM Conference on Hypertext and Hypermedia (2001):141–150

26. Yager, R.R.: Quantifier Guided Aggregation Using OWA Operators. International Journal of Intelligent Systems, 11 (1996): 49–73