A Collaborative Recommendation Framework for Ontology Evaluation and Reuse

Iván Cantador[1], Miriam Fernández[1], Pablo Castells[1]

Abstract. Ontology evaluation can be defined as assessing the quality and the adequacy of an ontology for being used in a specific context, for a specific goal. Although ontology reuse is being extensively addressed by the Semantic Web community, the lack of appropriate support tools and automatic techniques for the evaluation of certain ontology features are often a barrier for the implementation of successful ontology reuse methods. In this work, we describe the recommender module of CORE [5], a system for Collaborative Ontology Reuse and Evaluation. This module has been designed to confront the challenge of evaluating those ontology features that depend on human judgements and are by their nature, more difficult for machines to address. Taking advantage of collaborative filtering techniques, the system exploits the ontology ratings and evaluations provided by users to recommend the most suitable ontologies for a given domain. Thus, we claim two main contributions: the introduction of collaborative filtering notion like a new methodology for ontology evaluation and reuse, and a novel recommendation algorithm, which considers specific user requirements and restrictions instead of general user profiles or item-based similarity measures.

1 INTRODUCTION

The Semantic Web is envisioned as a new flexible and structured Web that takes advantage of explicit semantic information, understandable by machines, and therefore classifiable and suitable for sharing and reuse in a more efficient, effective and satisfactory way. In this vision, ontologies are proposed as the backbone technology to supply the required explicit semantic information.

Developing ontologies from scratch is a high-cost process that requires major engineering efforts, even when dealing with medium-scale ontologies. In order to properly face this problem, efficient ontology evaluation and reuse techniques and methodologies are needed. The lack of appropriate support tools and automatic measurement techniques for evaluating certain ontology features carries a shortage of information that is often a barrier for the successful of ontology reuse. Hence, in every day life we have to make choices considering incomplete information about the characteristics of the items that can be selected, and the whole set of available alternatives. It is in these situations, when we request our friends' knowledge and experience to be capable of taking the most appropriate decision. In this work, we shall exploit the benefits of the above natural social process to improve the actual approaches on ontology evaluation and reuse.

Specifically, we shall describe in detail the recommendation module of CORE [5], a Collaborative Ontology Reuse and Evaluation system. This tool provides automatic similarity measures for comparing a certain problem or Golden Standard to a set of available ontologies, and recommends not only those ontologies most similar to the domain of interest, but also the best rated ones by prior ontology users, according to several selected criteria.

[1] Universidad Autónoma de Madrid, Spain, emails: ivan.cantador@uam.es, miriam.fernandez@uam.es, pablo.castells@uam.es

The tool makes two main steps in the recommendation process. Firstly, it returns the ontologies most similar to the given Golden Standard. For similarity assessment, a user of CORE selects a subset from a list of comparison techniques provided by the system setting a number of standard ontology evaluation criteria to be applied. The system thus retrieves a ranked list of ontologies for each criterion. Afterwards, a unique ranking is defined by means of a global aggregated measure, which combines the different selected criteria using rank fusion techniques [2][12]

Secondly, once the system has retrieved those ontologies closely related to the Golden Standard, it performs a novel collaborative filtering strategy to evaluate and re-rank the considered ontologies. Since some ontology features can only be assessed by humans, this last evaluation step takes into consideration the manual feedback provided by users. Thus, the final ranked list will not only contain those ontologies that best fit the Golden Standard, but also the most qualified ones according to human evaluations.

The rest of the paper has the following structure. Section 2 summarizes relevant work related to our research. The system architecture is presented in Section 3, and our collaborative ontology recommendation algorithm is described in Section 4. Finally, some conclusions and future research lines are given in Section 5.

2 RELATED WORK

Our research addresses problems in different areas, where we draw from prior related work. In this paper we will focus our attention in two main topics: ontology evaluation and reuse, and collaborative filtering.

Different methodologies for ontology evaluation have been proposed in the literature considering the characteristics of the ontologies and the specific goals or tasks that the ontologies are intended for. An overview of ontology evaluation approaches is presented in [4], where four different categories are identified: those that evaluate an ontology by comparing it to a Golden Standard, or representation of the problem domain; those that evaluate the ontologies by plugging them in an application, and measuring the quality of the results that the application returns; those that evaluate ontologies by comparing them to unstructured or informal data (e.g. text documents) which represent the problem domain; and those based on human interaction to measure ontology features not recognizable by machines.

In each of the above approaches, a number of different evaluation levels might be considered to provide as much information as possible. Several levels can be identified in the literature: the lexical level, which compares the lexical entries of the ontology with a set of words that represent the problem domain; the taxonomy level, which considers the hierarchical connection between concepts using the is-a relation; the measurement of other semantic relations besides hierarchical ones; the syntactic level, which considers the syntactic requirements of the formal language used to describe the ontology; the context or application level, which considers the context of the ontology; and the structure, architecture and design levels which take into account the principles and criteria involved in the ontology construction itself.

On the other hand, collaborative filtering strategies [1][7][10][13] make automatic predictions (filter) about the interests of a user by collecting taste information from many users (collaborating). These predictions are specific to the user, differently to those given by more simple approaches that provide average scores for each item of interest; for example based on its number of votes.

Collaborative filtering is a widely explored field. Three main aspects typically distinguish the different techniques reported in the literature [9]: user profile representation and management, filtering method, and matching method.

User profile representation and management can be divided into five different tasks: *Profile representation* (accurate profiles are vital to ensure recommendations are appropriate and that users with similar profiles are in fact similar); *Initial profile generation* (the user is not usually willing to spend too much time in defining her/his interests to create a personal profile. Moreover, user interests may change dynamically over time); *Profile learning* (user profiles can be learned or updated using different sources of information that are potentially representative of user interests); *Profile adaptation* (techniques are needed to adapt the user profile to new interests and forget old ones as user interests evolve with time).

Filtering method. Products or actions are recommended to a user taking into account the available information (items and profiles). There are three main information filtering approaches for making recommendations: *Demographic filtering* (descriptions of people are used to learn the relationship between a single item and the type of people who like it); *Content-based filtering* (the user is recommended items based on descriptions of items previously evaluated by other users); *Collaborative filtering* (people with similar interests are matched and then recommendations are made).

Matching method. Defines how user interests and items are compared. Two main approaches can be identified: *User profile matching* (people with similar interests are matched before making recommendations); *User profile-item matching* (a direct comparison is made between the user profile and the items).

In CORE, a new ontology evaluation measure based on collaborative filtering is proposed, considering user's interest and previous human assessments of the ontologies.

3 SYSTEM OVERVIEW

In this section we describe the architecture of CORE, our Collaborative Ontology Reuse and Evaluation environment, focusing our attention on the collaborative recommender module. Figure 1 shows an overview of the system.

We distinguish three different modules. The first one, the left module in the figure, receives the Golden Standard definition as a set of initial terms, and allows the user to modify and extend it using WordNet [8]. This model of representation of the Golden Standard will not be explained here but could also be considered a novelty of our tool. The second one, represented in the centre of the figure, allows the user to select a set of ontology evaluation criteria provided by the system to recover the ontologies closest to the given Golden Standard. In this module we have introduced a novel lexical evaluation measure that exploits the semantic information stored in the Golden Standard model. The module also takes advantage of rank fusion techniques combining all the different evaluation criteria to obtain a final ontology raking. The third one, on the right of the figure, is the collaborative recommender module that re-ranks the list of recovered ontologies, taking into consideration previous feedback and evaluations of the users.

This module has been designed to confront the challenge of evaluating those ontology features that are by their nature, more difficult for machines to address. Where human judgment is required, the system will attempt to take advantage of collaborative filtering recommendation techniques [3][6][14]. Some approaches for ontology development [15] have been presented in the literature concerning collaboration techniques. However to our knowledge, collaborative filtering strategies have not yet been used in the context of ontology reuse.

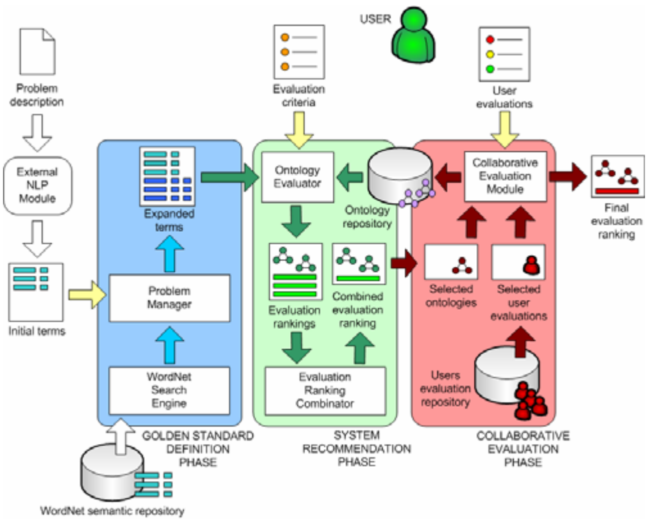

Figure 1. CORE architecture

Several issues have to be considered in a collaborative system. The first one is the representation of user profiles. The type of user profile selected for our system is a user-item rating matrix *(ontology evaluations based on specific criteria)***.** The initial profile is designed as a manual selection of five predefined criteria [11]:

- Correctness: specifies whether the information stored in the ontology is true, independently of the domain of interest.
- Readability: indicates the non-ambiguous interpretation of the meaning of the concept names.
- Flexibility: points out the adaptability or capability of the ontology to change.
- Level of Formality: highly informal, semi-informal, semi-formal, rigorously-formal.
- Type of model: upper-level (for ontologies describing general, domain-independent concepts), core-ontologies (for ontologies describing the most important concepts on a specific domain), domain-ontologies (for ontologies describing some domain of the world), task-ontologies (for ontologies describing generic types of tasks or activities) and application-ontologies (for ontologies describing some domain in an application-dependent manner).

The above criteria can be divided in two different groups: 1) the *numeric* criteria (correctness, readability and flexibility) that are represented by discrete integer values from 0 to 5, where 0 indicates the ontology does not fulfil the criterion, and 5 indicates the ontology completely satisfies it, and, 2) the *Boolean* criteria (level of formality and type of model) which are represented by a specific value that is either satisfied by the ontologies, or not. The collaborative module does not implement any profile learning or relevance feedback technique to update user profiles but, they can be modified manually.

After the user profile has been defined, it is important to select an appropriate type of filtering. For this work, a collaborative filtering technique has been chosen; this means, ontologies (our content items) are recommended based on previous user evaluations.

Finally, a matching strategy must also be selected for the recommendation process. In this work, a new technique for user profile-item matching is proposed. This novel algorithm will be explained in detail in section 4.

The right portion of Figure 2 shows the Collaborative Evaluation module. At the top level the user's interest can be selected as a subset of criteria with associated values, representing those thresholds that manual ontology evaluations should fulfil. For example, when a user sets a value of 3 for the correctness criterion, the system recognizes he is looking for ontologies whose correctness value is greater than or equal to 3. Once the user's interests have been defined, the set of manual evaluations stored in the system is used to compute which ontologies fit his interest best. The intermediate level shows the final ranked list of ontologies recommended by the module. To add new evaluations to the system, the user has to select an ontology from the list and choose one of the predetermined values for each of the five aforementioned criteria. The system also allows the user to add some comments to the ontology evaluation in order to provide more feedback.

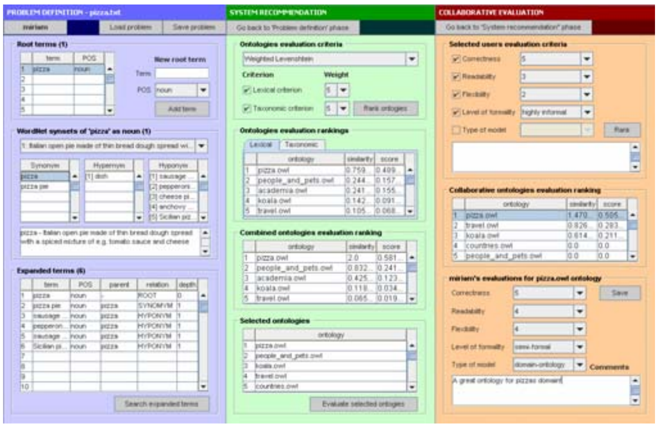

Figure 2. CORE graphical user interface

One more action has to be performed to visualize the evaluation results of a specific ontology. Figure 3 shows the user's evaluation module. On the left side, we can see the summary of the existing ontology evaluations with respect to the user's interests. In the figure, 3 of 6 evaluations of the ontology have fulfilled the correctness criteria, 5 of 6 evaluations have fulfilled the readability criteria, and so on. On the right side, we can see how the system enables the user to observe all the stored evaluations about a specific ontology. This might be of interest since we may trust some users more than others during the Collaborative Filtering process.

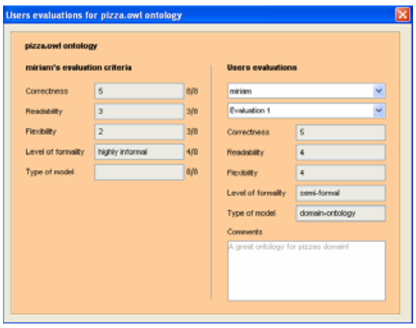

Figure 3. User evaluations in CORE

4 COLLABORATIVE ONTOLOGY EVALUATION AND RECOMMENDATION

In this section, we describe a novel ontology recommendation algorithm that exploits the advantages of collaborative filtering, and explores the manual evaluations stored in the system, for ranking the set of ontologies that best fulfils the user's interests.

As we explained in Section 3, user evaluations are represented as a set of five defined criteria and their respective values, manually determined by the users who made the evaluations. These criteria can have discrete numeric or non-numeric values. Moreover, user interests are expressed like a subset of the above criteria, and their respective values, meaning thresholds or restrictions to be satisfied by user evaluations.

Thus, a numeric criterion will be satisfied if an evaluation value is equal or greater than that expressed by its interest threshold, while a non-numeric criterion will be satisfied only when the evaluation is exactly the given "threshold" (i.e. in a Boolean or yes/no manner).

According to both types of user evaluation and interest criteria, *numeric* and *Boolean*, the recommendation algorithm will measure the degree in which each user restriction is satisfied by the evaluations, and will recommend a ranked ontology list according to similarity measures between the thresholds and the collaborative evaluations. Figure 4 shows all the previous definitions and ideas, locating them in the graphical interface of the system.

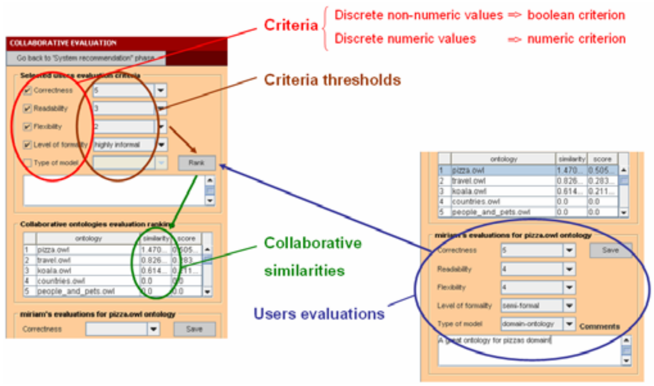

Figure 4. Two different types of user evaluation and interest criteria: *numeric* and *Boolean*

To create the final ranked ontology list the recommender module follows two phases. In the first one it calculates the similarity degrees between all the user evaluations and the specified user interest criteria thresholds. In the second one it combines the similarity measures of the evaluations, generating the overall rankings of the ontologies.

4.1 Similarity Measures for Collaborative Evaluation

In the current version of our system a user evaluate a specific ontology considering five different criteria (see Section 3). These five criteria can be divided in two different groups: 1) the *numeric* criteria (correctness, readability and flexibility), which take discrete numeric values [0, 1, 2, 3, 4, 5], where 0 means the ontology does not fulfil the criterion, and 5 means the ontology completely satisfy the criterion, and, 2) the *Boolean* criteria (level of formality and type of model), which are represented by specific non-numeric values that can be or not satisfied by the ontology.

Thus, user interests are defined as a subset of the above criteria, and their respective values representing the set of thresholds that should be reached by the ontologies.

Given a set of user interests, the system will size up all the stored evaluations, and will calculate their similarity measures. To explain these similarities we shall use a simple example of six different evaluations (E_1, E_2, E_3, E_4, E_5 and E_6) of a certain ontology. In the explanation we shall distinguish between the numeric and the Boolean criteria.

We start with the Boolean ones, assuming two different criteria, C_1 and C_2, with three possible values: "A", "B" and "C". In Table 1 we show the "threshold" values established by a user for these two criteria, "A" for C_1 and "B" for C_2, and the six evaluations stored in the system.

Table 1. Threshold and evaluation values for Boolean criteria C_1 and C_2

		Evaluations					
Criteria	Thresholds	E_1	E_2	E_3	E_4	E_5	E_6
C_1	"A"	"A"	"B"	"A"	"C"	"A"	"B"
C_2	"B"	"A"	"A"	"B"	"C"	"A"	"A"

In this case, because of the threshold of a criterion *n* is satisfied or not by a certain evaluation *m*, their corresponding similarity measure is simply 0 if they have the same value, and 2 otherwise.

$$similarity_{bool}(criterion_{mn}) = \begin{cases} 0 & \text{if } evaluation_{mn} \neq threshold_{mn} \\ 2 & \text{if } evaluation_{mn} = threshold_{mn} \end{cases}$$

The similarity results for the Boolean criteria of the example are shown in Table 2.

Table 2. Similarity values for Boolean criteria C_1 and C_2

		Evaluations					
Criteria	Thresholds	E_1	E_2	E_3	E_4	E_5	E_6
C_1	"A"	**2**	**0**	**2**	**0**	**2**	**0**
C_2	"B"	**0**	**0**	**2**	**0**	**0**	**0**

For the numeric criteria, the evaluations can overcome the thresholds to different degrees. Table 3 shows the thresholds established for criteria C_3, C_4 and C_5, and their six available evaluations. Note that E_1, E_2, E_3 and E_4 satisfy all the criteria, while E_5 and E_6 do not reach some of the corresponding thresholds.

Table 3. Threshold and evaluation values for numeric criteria C_3, C_4 and C_5

		Evaluations					
Criteria	Thresholds	E_1	E_2	E_3	E_4	E_5	E_6
C_3	≥ 3	3	4	5	5	2	0
C_4	≥ 0	0	1	4	5	0	0
C_5	≥ 5	5	5	5	5	4	0

In this case, the similarity measure has to take into account two different issues: the degree of satisfaction of the threshold, and the difficulty of achieving its value. Thus, the similarity between the value of criterion *n* in the evaluation *m*, and the threshold of interest is divided into two factors: 1) a similarity factor that considers whether the threshold is surpassed or not, and, 2) a penalty factor which penalizes those thresholds that are easier to be satisfied.

$$similarity_{num}(criterion_{mn}) =$$
$$= 1 + similarity^{*}_{num}(criterion_{mn}) \cdot penalty_{num}(threshold_{mn}) \in [0, 2]$$

This measure will also return values between 0 and 2. The idea of returning a similarity value between 0 and 2 is inspired on other collaborative matching measures [13] to not manage negative numbers, and facilitate, as we shall show in the next subsection, a coherent calculation of the final ontology rankings.

The similarity assessment is based on the distance between the value of the criterion *n* in the evaluation *m*, and the threshold indicated in the user's interests for that criterion. The more the value of the criterion *n* in evaluation *m* overcomes the threshold, the greater the similarity value shall be.

Specifically, following the expression below, if the difference *dif = (evaluation – threshold)* is equal or greater than 0, we assign a positive similarity in (0,1] that depends on the maximum difference *maxDif = (maxValue – threshold)* we can achieve with the given threshold; and else, if the difference *dif* is lower than 0, we give a negative similarity in [-1,0), punishing the distance of the value with the threshold.

$$similarity^{*}_{num}(criterion_{mn}) = \begin{cases} \dfrac{1 + dif}{1 + maxDif} \in (0,1] & \text{if } dif \geq 0 \\ \dfrac{dif}{threshold} \in [-1,0) & \text{if } dif < 0 \end{cases}$$

Table 4 summarizes the similarity* values for the three numeric criteria and the six evaluations of the example.

Table 4. Similarity* values for numeric criteria C_3, C_4 and C_5

		Evaluations					
Criteria	Thresholds	E_1	E_2	E_3	E_4	E_5	E_6
C_3	≥ 3	1/4	2/4	3/4	3/4	-1/3	-1
C_4	≥ 0	1/6	2/6	5/6	1	1/6	1/6
C_5	≥ 5	1	1	1	1	-1/5	-1

Comparing the evaluation values of Table 3 with the similarity values of Table 4, the reader may notice several important facts:

1. Evaluation E_4 satisfies criteria C_4 and C_5 with assessment values of 5. Applying the above expression, these criteria receive the same similarity of 1. However, criterion C_4 has a threshold of 0, and C_5 has a threshold equal to 5. As it is more difficult to satisfy the restriction imposed to C_5, this one should have a greater influence in the final ranking.
2. Evaluation E_6 gives an evaluation of 0 to criteria C_3 and C_5, not satisfying either of them and generating the same similarity value of -1. Again, because of their different thresholds, we should distinguish their corresponding relevance degrees in the rankings.

For these reasons, a threshold penalty factor is applied, reflecting how difficult it is to overcome the given thresholds. The more difficult to surpass a threshold, the lower the penalty value shall be.

$$penalty_{num}(threshold) = \frac{1 + threshold}{1 + maxValue} \in (0,1]$$

Table 5 shows the threshold penalty values for the three numeric criteria and the six evaluations of the example.

Table 5. Threshold penalty values for numeric criteria C_3, C_4 and C_5

Criteria	Thresholds	Evaluations					
		E_1	E_2	E_3	E_4	E_5	E_6
C_3	≥ 3	4/6	4/6	4/6	4/6	4/6	4/6
C_4	≥ 0	1/6	1/6	1/6	1/6	1/6	1/6
C_5	≥ 5	1	1	1	1	1	1

The similarity results for the numeric criteria of the example are shown in Table 6.

Table 6. Similarity values for numeric criteria C_3, C_4 and C_5

Criteria	Thresholds	Evaluations					
		E_1	E_2	E_3	E_4	E_5	E_6
C_3	≥ 3	**1.17**	**1.33**	**1.5**	**1.5**	**0.78**	**0.33**
C_4	≥ 0	**1.03**	**1.05**	**1.14**	**1.17**	**1.03**	**1.03**
C_5	≥ 5	**2**	**2**	**2**	**2**	**0.5**	**0**

As a preliminary approach, we calculate the similarity between an ontology evaluation and the user's requirements as the average of its *N* criteria similarities.

$$similarity(evaluation_m) = \frac{1}{N}\sum_{n=1}^{N} similarity(criterion_{mn})$$

A weighted average could be even more appropriate, and might make the collaborative recommender module more sophisticated and adjustable to user needs. This will be considered for a possible enhancement of the system in the continuation of our research.

4.2 Ontology Ranking and Recommendation

Once the similarities are calculated taking into account the user's interests and the evaluations stored in the system, a ranking is assigned to the ontologies.

The ranking of a specific ontology is measured as the average of its *M* evaluation similarities. Again, we do not consider different priorities in the evaluations of several users. We have planned to include in the system personalized user appreciations about the opinions of the rest of the users. Thus, for a certain user some evaluations will have more relevance than others, according to the users that made it.

$$\begin{aligned} ranking(ontology) &= \frac{1}{M}\sum_{m=1}^{M} similarity(evaluation_m) \\ &= \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} similarity(criterion_{mn}) \end{aligned}$$

Finally, in case of ties, the collaborative ranking mechanism sorts the ontologies taking into account not only the average similarity between the ontologies and the evaluations stored in the system, but also the total number of evaluations of each ontology, providing thus more relevance to those ontologies that have been rated more times.

$$\frac{M}{M_{total}} ranking(ontology)$$

5 CONCLUSIONS AND FUTURE WORK

We have presented CORE, a new tool for ontology evaluation and reuse, the main novel features of which include: a new Golden Standard model, new lexical evaluation criteria, the application of rank fusion techniques to combine different content ontology evaluation measures, and the use of a novel collaborative filtering strategy that takes advantage of user opinions in order to automatically evaluate features that only can be assessed by humans.

The collaborative module recommends those ontologies that best fit a certain problem domain, and have been best evaluated by the users of the system according to given specific evaluation criteria and restrictions. It is important to note here that although we have applied our recommendation method for assessing ontologies, it could be used in other very different applicative fields.

At the time of writing we are conducting initial experiments, not explained in this paper, that have been developed using a set of ontologies from the Protégé OWL repository[2]. The early results are clearly positive, but a more detailed and rigorously experimentation is needed in order to draw more conclusive and statistically significant observations.

REFERENCES

[1] Ardissono, L., et al. *INTRIGUE: personalized recommendation of tourist attractions for desktop and handset devices*. Applied Artificial Intelligence, Special Issue on Artificial Intelligence for Cultural Heritage and Digital Libraries 17(8-9), pp. 687-714, 2003.

[2] Aslam, J. A., Montague, M. *Models for metasearch*. 24th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'01). New Orleans, Louisiana, pp.276-284, 2001.

[3] Balabanovic, M., Shoham, Y. *Content-Based Collaborative Recommendation*. Communications ACM, pp. 66-72, 1997.

[4] Brank J., Grobelnik M., Mladenić D. *A survey of ontology evaluation techniques*. SIKDD 2005 at IS 2005, Ljubljana, Slovenia, 2005.

[5] Fernández, M., Cantador, I., Castells, P. *CORE: A Tool for Collaborative Ontology Reuse and Evaluation*. 4th Int. Workshop on Evaluation of Ontologies for the Web (EON 2006) at the 15th Int. World Wide Web Conference (WWW 2006). Edinburgh, UK, May 2006.

[6] Linden, G., Smith, B., York, J. *Amazon.com Recommendations: Item-to-Item Collab. Filtering*.IEEE Internet Computing, 7(1):76-80, 2003.

[7] McCarthy, J., Anagnost, T. *MusicFX: An arbiter of group preferences for computer supported collaborative workouts*. ACM International Conference on Computer Supported Cooperative Work (CSCW 1998). Seattle, Washington, pp. 363-372, 1998.

[8] Miller, G. *WordNet: A lexical database*. Communications of the ACM, 38(11): 39-41, 1995.

[9] Montaner M., López B., De la Rosa J.L. *A Taxonomy of Recommended Agents on the Internet*. AI Review 19: 285-330, 2003.

[10] O'Conner, M., Cosley, D., Konstan, J. A., Riedl, J. *PolyLens: A recommender system for groups of users*. 7th European Conference on CSCW (ECSCW 2001). Bonn, Germany, pp. 199-218, 2001.

[11] Paslaru, E. *Using Context Information to Improve Ontology Reuse*. Doctoral Workshop at the 17th Conference on Advanced Information Systems Engineering CAiSE'05, 2005.

[12] Renda, M. E., Straccia, U. *Web metasearch: rank vs. score based rank aggregation methods*. ACM symposium on Applied Computing. Melbourne, Florida, pp. 841-846, 2003.

[13] Resnick, P. et al. *GroupLens: An Open Architecture for Collaborative Filtering of Netnews*. Internal Research Report, MIT Center for Coordination Science, 1994.

[14] Sarwar, B.M., et al. *Item-Based Collaborative Filtering Recommendation Algorithms*. 10th International World Wide Web Conference, ACM Press, pp. 285-295, 2001.

[15] Sure, Y., et al. *OntoEdit: Collaborative Ontology Development for the Semantic Web*. ISWC 2002.

[2] http://protege.stanford.edu/plugins/owl/owl-library/index.html