

Mining user competency from semantic trace

Ning Wang, Marie-Hélène Abel, Jean-Paul Barthès, Elsa Negre

► **To cite this version:**

Ning Wang, Marie-Hélène Abel, Jean-Paul Barthès, Elsa Negre. Mining user competency from semantic trace. Computer Supported Cooperative Work in Design (CSCWD), May 2015, Calabria, Italy. 10.1109/CSCWD.2015.7230932 . hal-01266493

HAL Id: hal-01266493

<https://hal.archives-ouvertes.fr/hal-01266493>

Submitted on 3 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mining User Competency from Semantic Trace

Ning Wang, Marie-Hélène Abel, Jean-Paul Barthès
UMR CNRS 7253 HEUDIASYC
University of Technology of Compiègne
Compiègne, France
{ning.wang, marie-helene.abel, barthes}@utc.fr

Elsa Negre
UMR CNRS 7243 LAMSADE
Université Paris Dauphine
Paris, France
elsa.negre@dauphine.fr

Abstract—In order to achieve individual or collective goals, users in informational environments collaborate to integrate intellectual resources and knowledge. Thanks to informational environments, users can better organize, realize and record collaboration. Every activity produces a set of traces. Such traces can be recorded and classified, based on a model of traces. With the help of a model of competency, these traces also contribute to evaluate the competency of users on certain subjects. In this article, we propose a semantic model of traces and analyze classified traces by means of TF-IDF. We also considered the impact of time on the decreasing importance of traces. We show how to offer users recommendations and decision aid.

Keywords—traces of interaction; competency; semantic modelling; decision aid; recommender system;

I. INTRODUCTION

Grudin proposes that collaboration is a source of power for human society development and progress [1]. As the complexity and difficulty of our projects increase, the need for collaboration increases. Thanks to information technology, nowadays collaboration is more and more organized in informational environment. In such an environment, users achieve their purposes by taking different actions. We are interested in the results of actions as well as the actions themselves. For example, we care not only about the content of a document a user has created, but also in knowing who shares and consults this document at what time, with what frequency. A set of actions, step by step, is defined as a trace by Zarka et al. [2]. Under careful modelling and analysis, traces could in return help indicating the competency of an individual or a group of users as shown by Tomaz et al. [3]. Thus, with the information exploited behind traces, we can improve collaboration, as mentioned by Garcia-Crespo et al. focusing on the reuse of traces for different purposes such as decision aid [4] or by Chang et al. for recommendation [5].

In this paper we propose a mechanism that models, records and analyzes users' traces. It allows evaluating competencies for recommending users with more expertise on certain subjects for either personal purpose or to an entire group of users. The following tasks are needed to achieve this objective: (i) propose a semantic structure to record traces; (ii) evaluate traces using TF-IDF [6] and a semantic distance among the actions which compose traces; (iii) propose a model of

competency which integrates our methods mentioned above; and (iv) propose recommendations and provide some decision aid accordingly.

The remainder of this paper is organized as follows. We identify various limitations of the current studies on recommendation in Section 2. Then we propose a kind of recommender system for the need of exploiting traces. In Section 4 we illustrate our method by giving a toy example. Section 5 gives conclusions and points to directions for future works.

II. RELATED WORKS

Aiming at a better treatment and exploitation of traces, we need to analyze different types of traces as well as the structure of collaboration in a Collaborative Working Environment (CWE). The model of traces proposed by Li et al. [7] allows an elaborated analysis of interactions among users. It pays special attention to the exchanges amongst users in informational environment. Then traces can be exploited to feed a recommender system. The interests of a recommender system is justified by the need to manage the growing amount of information as stressed by Admavicius and Tuzhilin [8].

Recently, various articles were published aiming at exploiting the traces with the help of semantics. Chen et al. [9] presents a mechanism for personalized knowledge search and recommendation adapting a suitable domain ontology according to the previous browsing and reading behavior of users. Vesin et al. [10] presents a new approach for performing effective personalization based on semantic web technologies. [11], [12] and [13] all tried to provide recommender using probabilistic modeling. But these studies do not work much on modelling users' actions although a good model of organizing and storing users' actions can make future exploitation of traces more efficient.

III. OUR APPROACH

Fig. 1 shows the structure of our proposed recommender system. First, with the model of actions we propose, users' actions are collected and modelled from an interactive platform. After being sifted by the filter of classification, we obtain classified traces, which allow a preliminary presentation back to the users. Alternatively, we apply an algorithm to

calculate an index indicating the correlation between the classified traces of a certain user and a given subject. These values can lead to useful information that are presented as personalized recommendations, either to a group defined as a set of users of the platform, or to an individual user.

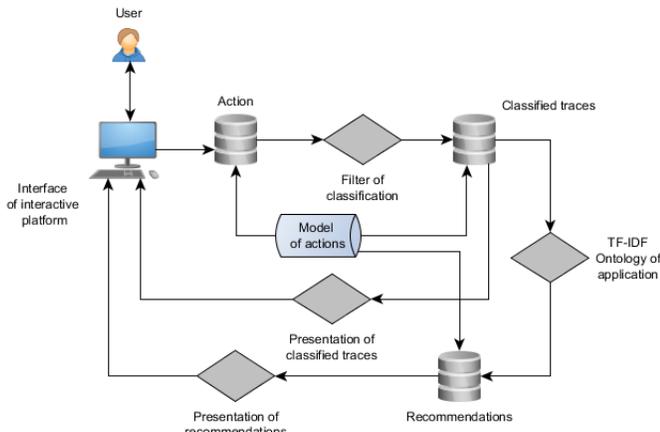


Fig. 1. The structure of our proposed recommender system for the exploitation of semantic traces.

A. Modelling of Actions

We define the principal concepts as follows:

- *Action*: an interaction or an act performed by a user in a collaborative environment, e.g. sending a document to other users.
- *Classified trace*: a set of actions that were performed by a user in the informational environment classified

according to the model of traces [7].

- *Set of traces*: an ordered set of classified traces.

According to our definition, action is the most basic element forming a trace. Regarded as an important resource for our recommender system, we introduce the Resource Description Framework (RDF) to model actions [14]. RDF is used as a general formalism for conceptual description or modelling of information that is implemented in web resources. Fig. 2 shows the basic structure in the RDF schema of our model. An ellipse represents a class of resources and a rectangle represents an object property. For example, a person has the object property “has_id_person” and the range of this property is a class called “id”. In Fig. 3, the class “creation” inherits all properties from “action”.

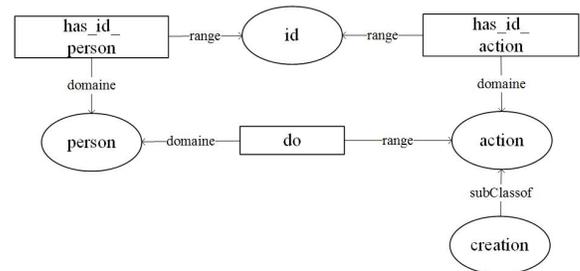


Fig. 2. Basic structure in RDFS graph presenting an action.

We apply the model of actions to a web-based collaborative platform E-MEMORAE2.0 proposed by Abel and Leblanc [15]. Details are shown in Fig. 3. This model of actions has two main advantages compared to a traditional form of history or log of users:

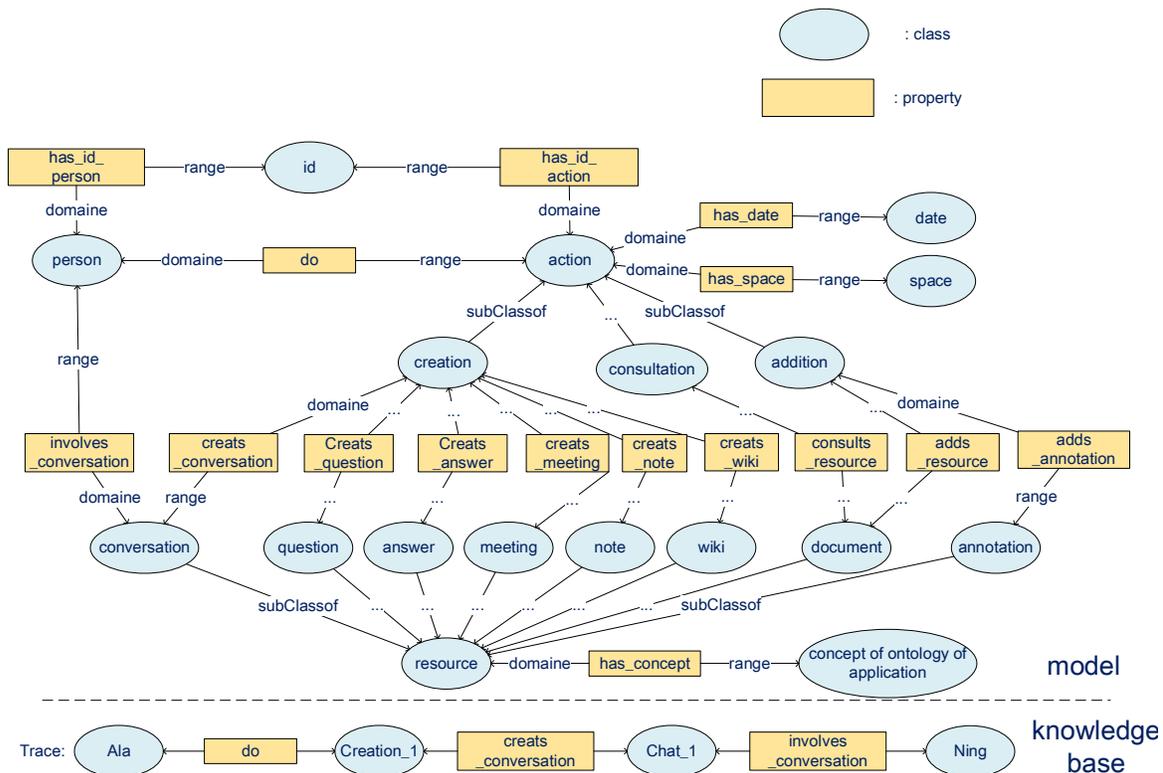


Fig. 3. The model of actions in the platform E-Memora 2.0.

- Actions are presented in a labeled, directed multi-graph. In our model the actions are represented as resources in the RDF schema and they are connected by properties. This allows a better structure of storage and usage of actions. For example, a person “Ala” chats with “Ning”. This action can be presented by an RDF instance showed in the lower part of Fig. 3 where “Ala” and “Ning” are two instances of the class of resource “person”. “Chat_1” is an instance of the class of resource “conversation” which is linked to the action “creation”.
- Normally different types of actions have different importance. For example creating elements of a Wiki is more important than consulting it. In our model actions are classified by three classes: creation, consultation and addition which enables us to treat different types of actions more efficiently. We define the importance of actions more efficiently. We define the importance of creation, addition and consultation respectively as “high”, “medium” or “low”.

B. Application of TF-IDF

We apply TF-IDF to evaluate the importance of traces. TF-IDF (Term Frequency-Inverse Document Frequency) [6] is a statistical method used for the research of information in textual documents. It measures the correlation of a term appearing in a document from a corpus. The weight of a term is proportionally heavier when it appears more frequently in a document. It also varies with the frequency of the word in the corpus. The values are used for evaluating the importance of a document.

Here we adapt this method to our case. Fig. 4 illustrates the key-concepts of TF-IDF and those relevant in our case. Typically TF-IDF focuses on the relation between words, documents and corpus. If a word appears more in a document and at the same time appears less in the other documents of the same corpus, it better represents this document. For our research, we are interested in evaluating the correlation between a trace of a given user and a certain subject. We propose to consider that if the actions of a user are more pertinent concerning a subject, then the user has more knowledge about it. So we are able to recommend this user as an expert in this concept. In our case, we study the relation between actions, traces and the set of traces in a group of users working on the same subject. As mentioned in the section III.A, we distinguish the importance of different types of actions as “high”, “medium” or “low” (Table 1).

TABLE I. THE WEIGHT OF DIFFERENT TYPES OF ACTIONS.

Action importance	Description	Action type	Weight
High	Users are highly aware of the concept	Creation	1.0
Medium	Users have contribution to this concept	Addition	0.8
Low	Users need knowledge about the concept	Consultation	0.6

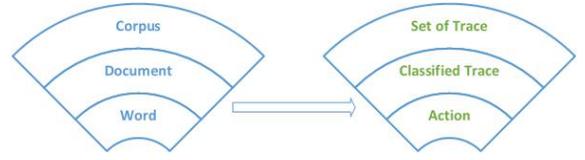


Fig. 4. Relation between concepts of the method TF-IDF and the method adapted to our case.

We adapt the equation of TF-IDF:

$$tf_{i,j} = \frac{\sum_i n_{i,j} \cdot w_i}{\sum_k n_{k,j} \cdot w_k} \quad (1)$$

$$idf_i = \log \frac{|P|}{|\{p : t_i \in p\}|} \quad (2)$$

Where:

- $\sum_i n_{i,j} \cdot w_i$: the sum of weights of actions concerning subject i performed by user j .
- $\sum_k n_{k,j} \cdot w_k$: the sum of weights of actions concerning all k subjects performed by user j .
- $|P|$: the number of users in a group.
- $|\{p : t_i \in p\}|$: the number of users in a group who have performed at least one action about subject i .

The value of TF-IDF, $C_{i,j}$ indicating the competency of user j on subject i , which can be regarded as the relevance between a subject and a user, is defined as follows:

$$C_{i,j} = tf_{i,j} \times idf_i \quad (3)$$

C. Application of Semantics

When we evaluate the competency of a user on a certain subject, we can also take into consideration the semantic relationship among concepts. As semantics indicates the real meaning, a smaller semantic distance between two concepts means a closer relation between them. For instance, if we need to find an expert on "Java", even though a person hasn't left a weighted trace on "Java" directly, he/she may have contributed to a project which requires the ability of "Java" in a semantic view. So we can evaluate the weight of this user on this certain project. We will explain in Section 4 by giving a toy example.

D. Classified Recommender System

We take the model of traces proposed by Li et al. [7] as the basis to carry out classified recommendations. According to this model, a trace is classified into 4 types: private trace, collaborative trace, collective trace and personal trace. A private trace is sent and received by the same user. A collaborative trace has one sender and at least two receivers. A collective trace has many senders and receivers while a personal trace is defined as having only one sender with no limit on its receivers.

Our recommender system takes into consideration a semantic model of the system along with the traces of recorded interactions. It aims at realizing recommendations for a group (improve the collaboration, identify risks, opportunities of a set of users from a group), of an individual (how to improve his efficiency, the organization of a user among his tasks), for private purposes (how to improve the private organization of a user) and for collective purposes (how to improve the communication inside a group, etc.).

E. Decreasing Importance of a Trace in Time

In a collaborative environment, the date when an action is carried out is also recorded. The forgetting curve hypothesizes the decline of memory retention in time. This curve shows how information is lost over time when there is no attempt to retain it as mentioned by Averell and Heathcote [16]. A typical graph of the forgetting curve purports to show that humans tend to halve their memory of newly learned knowledge in a matter of days or weeks unless they consciously review the learned material. So, it is necessary to take the decay of knowledge with time into consideration. In general, a recent action has more weight than a previous action. One of the common methods is to apply the impact of time on the importance of trace by a decay function. It is often used in time-sensitive recommender system claimed by Koren [17], Zhang and Liu [18], or Chen and Zhao [19].

$$f(t) = e^{-\lambda t} \quad (4)$$

The decay function (4) indicates that the trace we observe and analyze is less and less important as time goes by. Its importance decays fast at first and then decays more slowly. The parameter λ helps adjust the speed of decay by t . Baugh indicates this regular pattern fits better the psychological pattern of humanity [20]. Fig.5 shows the image of decay function. Suppose the importance of trace decreases to 50% of its original value after 5 weeks and correspondingly $\lambda = 0.15$.

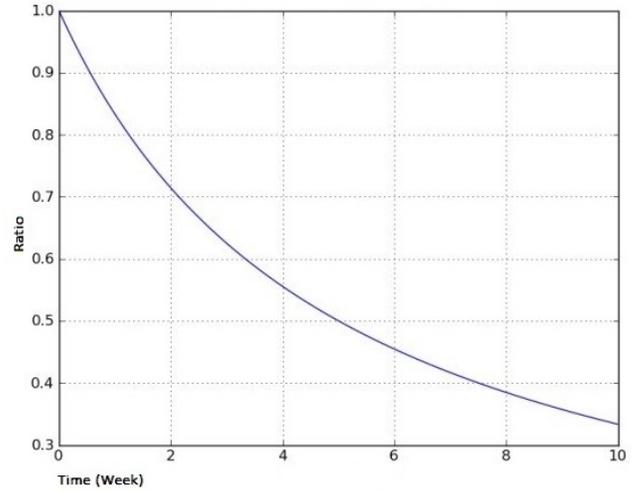


Fig. 5. Importance decay by time in week.

F. Model of Competency

A competency is the capability to apply a set of related knowledge, skills, and abilities to successfully perform functions or tasks in a defined work setting. In this section we orchestrate the methods above to compose the model of competency. According to Baugh [20], competency is distinguished into two categories:

- Hard competency, which identifies the intellectual process during the realization of activities.
- Soft competency, which corresponds to personal behavior, e.g., aptitude to work with others and leadership, etc.

Our work focuses on hard competency. In the following, for reasons of conciseness, we simply use competency instead of hard competency.

In recent years, many works have focused on the modelling of competency. Molina and Flores [21] defines a “Core Competencies” in the manufacturing clusters, including 4 generic and comprehensive components, namely: products, processes, skills, and task service. Müller [22] proposes “Competence cells” for the competency cell-based networks in which the main components are “resources” and “fulfilled task or executed function”. Boucher et al. [23] presents a “s-a-r-C model” of competency consisting of “Professional Situations”, “Actor” and “Resource”. These models share two components, namely “resources” (including “human resources” or “physical resources”) and “activity” (also called “process”, “production skill” or “task”).

The success of an activity requires actions on relevant concepts. For example, if we want to create a website, collaborators should put their knowledge about different concepts such as “PHP”, “Javascript” and “HTML” into effort. During the activity, their actions are recorded, e.g., creating a manual or sharing a technical article. Our system can analyze these actions and evaluate collaborators’ competencies on different subjects, so that when the next time arrives, and a certain activity needs an expert on a certain concept, we can

recommend a collaborator. As our system aims at the management of knowledge, resource is presented mainly in the form of knowledge. Thus we propose an “action-knowledge model” that integrates the merits in the models above and covers the strength of our system as showed in Fig. 6.

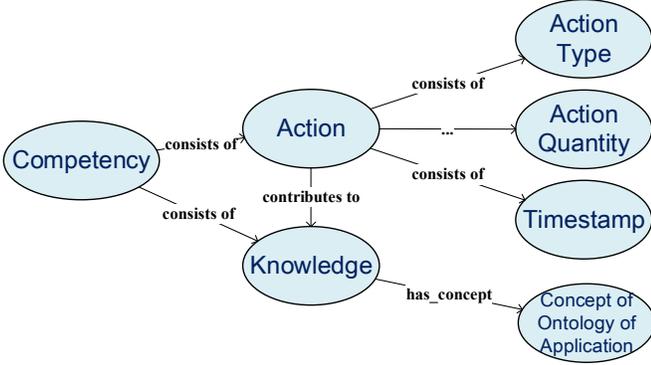


Fig. 6. Generic a-k model of competency

Detailed definition of these components is as following:

- Action is how a user applies the knowledge. Action also helps to accumulate the knowledge of a user. For example, if a user consults many files about “Java”, his/her knowledge grows.
- Action Type describes different types of actions. Some types of actions directly contribute to competency, for example answering questions of other users or creating a Wiki about this concept. These actions indicate that the user tends to be more competent about what he/she applies. Meanwhile other actions only contribute to the knowledge of the concept such as reading a piece of paper about it. In section III.A, action types are described by semantic model of actions.
- Action Quantity records both a user and his collaborators’ intensity of efforts on this activity as mentioned in section III.B.
- Timestamp records the time when an action takes place, as mentioned in section III.E.
- Knowledge is what a user applies during an action.
- Concept of Ontology of Application describes the nature of a user’s action. It is the semantic description of knowledge, as presented in section III.C.

With this model of competency we merge our methods in this section to evaluate users’ competencies.

IV. EXAMPLE

In order to make a recommendation to a user that tends to have more expertise, we use our recommender system into the collaborative working environment. Fig. 7 illustrates an example of collaboration on different subjects of two groups of users with a histogram chart. Each line represents the collaborative trace of a user on different subjects.

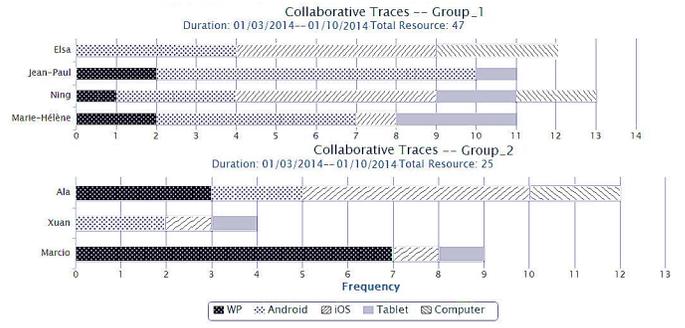


Fig. 7. An example of collaborative interactions of two groups.

Now we evaluate the competency of user “Ning” on “WP”. According to Fig. 7, “Ning” has realized 13 actions of which 1 action concerns “WP”. We suppose these actions happened 3 weeks ago. In group 1, the number of users is 4 of which 3 have realized at least one action on “WP”. Details of traces are showed in Table II.

TABLE II. TRACES OF ALA AND NING CONCERNING “WP”

User	Action type	Weight	Happened week(s) ago
Ning	Creation	1.0	3
Ala	Addition	0.8	13
	Addition	0.8	14
	Consultation	0.6	16

According to (3) and (4), we obtain:

$$C_{WP,Ning} = \frac{1}{13} \times \log \frac{4}{3} \times e^{-0.15 \times 3} \times 1.0 = 6.128 \times 10^{-3} \quad (5)$$

Likewise, we obtain the index of user Ala on subject “WP” : $C_{WP,Ala} = 3.906 \times 10^{-3}$. With the traces of two users and other users in the same group, we can conclude that “Ala” is less competent than “Ning” on “WP” relying on $C_{WP,Ala} < C_{WP,Ning}$.

Meanwhile, if we lack direct relevant traces on a concept, we turn to making a complex request on the semantic concepts connected to it. Fig. 8 shows a part of the domain ontology of a project, the users involved and the relevant competencies required. It shows that four users are members of “Group_C” working on “Project_M”. This group owns “Sharing_Space_S” where all actions take place. Moreover, this project requires competencies such as “Java”, “PHP”, and “C#”. Now we tend to evaluate different users’ competencies on a concept “Java”. If no trace is recorded between “Java” and an individual user, we make a complex request on all concepts semantically related to “Java”. Finally we find that the concept “Project_M” is ideal as it requires competencies including “Java”. So we evaluate all participants on this project and the results are shown in Table III. After comparison we find that user “Ala” is most involved in this project. So, we believe that he is more competent on “Java”.

TABLE III. THE INDEX OF COMPETENCY OF USERS ON SUBJECT "PROJECT_M".

User	C(Project_M, user)
Ning	0.645
Marcio	0
Ala	3.224
Hanen	1.290

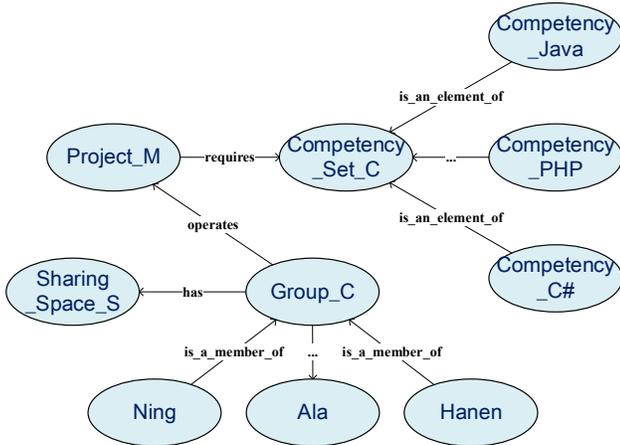


Fig. 8. Semantic relationship of a project, the users involved and the relevant competencies required.

V. CONCLUSION AND FUTURE WORK

A fully exploitation of traces helps us organizing and improving collaboration. In this article we proposed recommendations based on the evaluation of traces using TF-IDF with the consideration of time decay of knowledge. Moreover we demonstrated how to solve the problem when there does not exist enough relevant traces with semantics. Finally, we orchestrated these methods to evaluate the competencies of users. We illustrated our method by an example.

Future work includes implementing our proposal of recommender system. Moreover, different types of resources should be measured separately to obtain a complete model of competency.

REFERENCES

- [1] Grudin, J. (1994). Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM*, 37(1), 92-105.
- [2] Zarka, R., Cordier, A., Egyed-Zsigmond, E., & Mille, A. (2011). Trace replay with change propagation impact in client/server applications. *Actes de IC2011*, 607-622.
- [3] Tomaz, L. F., Nt, J. A. R., Souza, J. M., & Xexéo, G. B. (2011, June). Bringing knowledge into recommendation systems. In *Computer Supported Cooperative Work in Design (CSCWD)*, 2011 15th International Conference on (pp. 246-252). IEEE.
- [4] García-Crespo, Á., López-Cuadrado, J. L., Colomo-Palacios, R., González-Carrasco, I., & Ruiz-Mezcua, B. (2011). Sem-Fit: A semantic

based expert system to provide recommendations in the tourism domain. *Expert systems with applications*, 38(10), 13310-13319.

- [5] Chang, N., Irvan, M., & Terano, T. (2013). A TV Program Recommender Framework. *Procedia Computer Science*, 22, 561-570.
- [6] Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1), 11-21.
- [7] Li, Q., Abel, M. H., & Barthès, J. P. (2012). Facilitating Experience Sharing Groups-Collaborative Trace Reuse and Exploitation. In *Proceeding of International Conference on Knowledge Management and Information Sharing* (pp. 21-30).
- [8] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6), 734-749.
- [9] Chen, Y. J., Chu, H. C., Chen, Y. M., & Chao, C. Y. (2013). Adapting domain ontology for personalized knowledge search and recommendation. *Information & Management*, 50(6), 285-303.
- [10] Vesin, B., Ivanović, M., Klačnja-Milićević, A., & Budimac, Z. (2012). Protus 2.0: Ontology-based semantic recommendation in programming tutoring system. *Expert Systems with Applications*, 39(15), 12229-12246.
- [11] Breese, J. S., Heckerman, D., & Kadie, C. (1998, July). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (pp. 43-52). Morgan Kaufmann Publishers Inc..
- [12] Condliff, M. K., Lewis, D. D., Madigan, D., & Posse, C. (1999, August). Bayesian mixed-effects models for recommender systems. In *ACM SIGIR (Vol. 99, pp. 23-30)*.
- [13] Pennock, D. M., Horvitz, E., Lawrence, S., & Giles, C. L. (2000, June). Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence* (pp. 473-480). Morgan Kaufmann Publishers Inc..
- [14] Antoniou, G., & Van Harmelen, F. (2004). *A semantic web primer* (pp. 253-260). MIT press.
- [15] Abel, M. H., & Leblanc, A. (2009, October). Knowledge sharing via the E-MEMORAe2. 0 platform. In *Proc of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning* (pp. 10-19).
- [16] Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, 55(1), 25-35.
- [17] Koren, Y. (2010). Collaborative filtering with temporal dynamics. *Communications of the ACM*, 53(4), 89-97.
- [18] Zhang, Y., & Liu, Y. (2010, April). A Collaborative filtering algorithm based on time period partition. In *Intelligent Information Technology and Security Informatics (IITS)*, 2010 Third International Symposium on (pp. 777-780). IEEE.
- [19] Chen Z M, Jiang Y, Zhao Y. A Collaborative Filtering Recommendation Algorithm Based on User Interest Change And Trust Evaluation[J]. *International Journal of Digital Content Technology and its applications*, 2010, 9(4): 106-113.
- [20] Baugh, Jeremy, "Rewarding Competencies in Flatter Organizations," *Competency: the journal of performance through people*, v. 4:3, Spring, 1997
- [21] Molina, A., & Flores, M. (1999). A virtual enterprise in Mexico: From concepts to practice. *Journal of Intelligent and Robotic Systems*, 26(3-4), 289-302.
- [22] Müller, E. (2006). Production planning and operation in competence-cell-based networks. *Production Planning & Control*, 17(2), 99-112.
- [23] Boucher, X., Peillon, S., & Burlat, P. (2005). Towards a decision support for a collaborative increase of competencies within networks of firms. *Ecole de Mines de Saint-Etienne, Research Report G21-EMSE*, 600-009.