

# Methods for Collective Intelligence Utilization in Distributed Knowledge System

D. Rutkauskiene, E. Karazinas

*E-Learning Technology Centre, Kaunas University of Technology,  
Studentu St. 48A, LT-51367 Kaunas, Lithuania  
danguole.rutkauskiene@ktu.lt*

**Abstract**—The aim of this paper is to present requirements and perform the analysis of methods for grouping social networking system users based on their digital artifacts in the field of R&D by analyzing e-portfolios with knowledge discovery algorithms. In this paper we present the system architecture to collect and maintain conference presentation recordings, papers and other research activity indicators in user's personal e-portfolio. Distributed learning environment is used for digital e-portfolio artifact gathering, it consists of open source learning tools for conferences organization, virtual video presentations and e-journal systems. Analysis of the methods is carried out, which utilize collective intelligence and enable recommendation services to support scientists' collaboration by bringing them together, revealing intersected fields of interests for joint research empowerment, recommending conference presentations, publications, participation in joint projects. Systems architecture is proposed together with supplementary tools, that are provided to sustain science community's cooperation in discussion groups, virtual meetings and joint publications.

**Index Terms**—Distributed databases, information management, information processing, distributed information systems.

## I. INTRODUCTION

Web 2.0 internet empowers people to collaborate in new ways, which were not possible before [1]. Information, virtual media, virtual collaboration gradually comes into real world through social networking systems and will embed itself seamlessly as internet technology evolves. Therefore effective collaboration methods must be adapted, to take full advantage of new technological possibilities. By fostering joint academic community activities closer cooperation can be achieved, that would result in fruitful scientific and economic results. Students, lecturers and researchers would exchange experience in more effective ways and go into new insights if systems could support collaboration process by automating the gathering process of research results information into databases and based on effective methods make recommendations for joint work whereas additional tools would support joint efforts. A project for developing such system is started.

Manuscript received February 28, 2012; accepted June 15, 2012.

This paper has been supported by EUREKA scientific programme ITEA 2 Call 4 09020 (European Structural Funds project No. VP1-3.1-ŠMM-06-V-01-003, VICAM).

## II. VICAMC – VIRTUAL CONFERENCES AND MEETINGS CENTER

VICAMC Project aims to develop a distributed platform that would enable to migrate all aspects of conferences or meetings to virtual environment as well as to enhance traditional events with innovative collaboration, content authoring, knowledge sharing and semantic web technologies. Scenarios for the physical meetings such as conferences, seminars, symposiums, workshops or exhibitions has evolved during decades and proved to be effective for communication among different groups of people.

Rapid development of information technologies and broadband Internet services creates new possibilities to communicate on-line and to transform physical meetings in many ways. For example, video conferences has been used for few decades as an alternative for physical meetings, Internet broadcasts and on demand video allowed to extend auditorium of participants dramatically, on-line collaboration tools made it possible to work on joint projects and to collaborate while authoring common paper or any other digital content, and so on. On the other hand there are many systems developed for accommodation of physical event organizational issues such as user registration, paper submission, peer-review process, etc.

In this situation we have a problem of scattered information across various repositories and different tools. To create new possibilities to communicate and share information on the basis of synergetic holistic approach, there is a need for a platform, which would offer integration services for separate repositories and could allow creation of new services on top of them. Additionally, it would offer easy tools for managing data and users from one location.

The idea of the project is to go beyond capabilities of event management system or separate on-line collaboration tools and to create a platform that would provide services to both event organizers and individuals:

- 1) *For event organizer system would enable to extend their audience from physical to virtual participants or even to move whole event on-line;*
- 2) *For authors system would allow to use on-line collaboration and presentation tools both for virtual and physical events by additionally extending audience of author's listeners and building his on-line community of*

interests by collecting all digital artifacts authored by him in one virtual shared portfolio;

3) For all participants system would allow to take part in live events as well as to search and watch for recorded presentations.

The integration between those services would create new possibilities to get additional benefit not only from separate, but also from overall collected and integrated digital assets - through relations, where different type of media, generated and connected together through the same event would represent some kind of Unit of Knowledge that would support knowledge sharing and promote its higher quality:

1) The discovery of information would be enhanced through complementary meta-data introduced to different digital assets;

2) Each digital asset in this Unit of Knowledge would be supported by other assets and would form the pool of interconnected description data.

Going further, this pool of interconnected digital assets could be expanded by adding related information from other Units of Knowledge, connecting other media types as well as user generated data from notes and subject related discussions.

All digital artifacts collected on the platform would allow maximum interaction possibilities with other internet databases and learning object repositories, and would facilitate development of subject oriented virtual communities. Each member of these communities would have its own personal information profile describing his field of interest. Automated interest mapping would instantly connect different users on-line by providing convenient facilities for virtual meetings and other interaction possibilities as well as by promoting green meetings against physical traveling.

### III. ARCHITECTURE OF THE SYSTEM

VICAMC distributed learning environment architecture consist of partners and one central installation (Fig. 1). Distributed environment was chosen because of workloads for workstations of video presentations and the need to customize conference websites. Partner installation tools:

1) *Drupal* [2] CMS – a comprising system for user management, that has modules:

- Video Presentation module;
- E-journal module;

2) *Conference Organization Distribution* [3] installation profile with additional customizations for project needs.

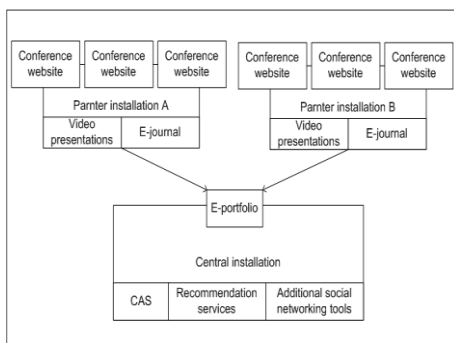


Fig. 1. VICAMC System architecture.

Central installation tools:

1) *Drupal CMS* – a comprising system for user management, that has modules:

- CAS (Central Authentication Service) – for user management and SSO solution to partner installations;
- E-portfolio module;
- Discussion forums;
- Recommendation services.

Digital participants' artifacts from partners' installations are collected to the central installation personal e-portfolio. CAS service enables participants to login to any of partner installation, no matter where the conference takes place. Users only need to register one time at one of partner installations and participate in the conference. After registration they also can login to central installation and use recommendation services or participate in social networking with other colleagues, which were met in the conferences.

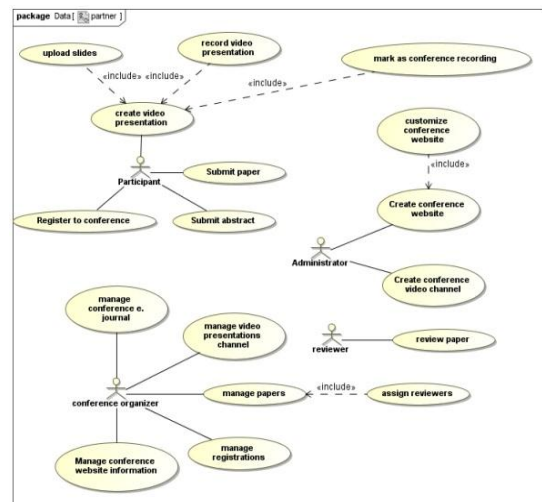


Fig. 2. Use cases for partner installation.

In Fig. 2 use cases are depicted of different user roles for partner installation:

1) *Administrators* – create new conference website, customize it to the conference organizers needs, create conference video channel;

2) *Participants* may register to the conference through conference website, they also can submit abstract and full papers. When conference takes place, they can create video presentation, upload slides and make a recording;

3) *Conferences organizers* can manage registrations, papers, assign reviewers, edit other important information about conference, manage video presentation channel and manage conference e. journal;

4) *Reviewers* review assigned papers.

In Fig. 3 use cases for central installation are depicted of different user roles:

1) *Participants* may browse video presentations, that are collected from partners, browse e. journal papers, revise personal e. portfolio and add additional remarks, send invitations to specific participants to initiate new project, use social networking tools – video discussions, text forums, chats to collaborate with other participants. Participants also can get recommendations that may be

interesting – video presentations, other participants, papers;

2) Conference organizers may announce new conferences to specific participants, which may be interested to participate.

For these use cases to be realized, information to personal e-portfolios must be collected and methods for recommendation services must be implemented.

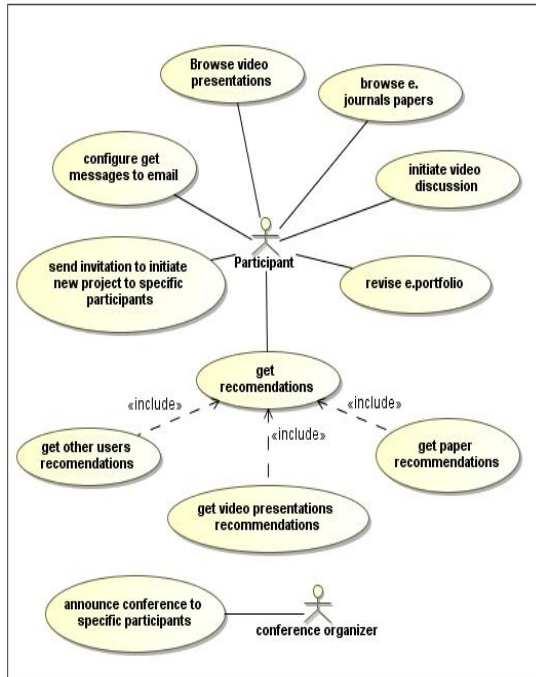


Fig. 3. Use cases for central installation.

There are digital elements and research activity indicators, which compose user's e-portfolio:

- 1) Registration to the conference;
- 2) Registration with abstract;
- 3) Registration with publication;
- 4) Publication, that was accepted and published in e-journal;
- 5) Video presentation's slides;
- 6) Viewed video presentations;
- 7) Viewed publications;
- 8) Searches performed;
- 9) Participation in discussion groups;
- 10) Video and publication ratings.

All these artifacts can be collected and used to enable recommendation services, but appropriate methods for collective intelligence utilization must be selected.

#### IV. ANALYSIS OF RECOMMENDATION SERVICES ENABLING METHODS

The system will analyze e-portfolios and will create groups of interests. Services, which will use these groups of interests, will be able:

- 1) To send an invitation to participate in the project for users, that will be recommended;
- 2) To recommend discussion groups, that one might be interested;
- 3) To get recommendation to collaborate with other persons in the system;

4) To get content based recommendations;

5) Video presentations;

6) E-journal publications;

7) To annotate and invite users to new conferences;

8) To disseminate information about results in other projects.

To enable these services different e-portfolio analysis methods can be used.

#### V. USER-BASED COLLABORATIVE FILTERING

Collaborative filtering (CF) is a technique used by some recommender systems. Currently, collaborative filtering (CF) is the most commonly used and studied technology [4], [5]. Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). The collaborative filtering algorithms that use similarities among users are called user based collaborative filtering [6], [7]. The underlying assumption of the collaborative filtering approach is that if a person A has the same opinion as a person B on an issue, A is more likely to have B's opinion on a different issue x than to have the opinion on x of a person chosen randomly. By this filtering method weights are assigned to users based on similarities of their ratings with that of the target user [8].

For calculating the similarity between a one user and another user, different similarity metrics can be used:

- 1) Cosine similarity;
- 2) Pearson correlation;
- 3) Jaccard-Tanimoto index;
- 4) Sorensen coefficient.

##### A. Cosine similarity

Cosine similarity is a measure of similarity between two vectors by measuring the cosine of the angle between them. The cosine of 0 is 1, and less than 1 for any other angle; the lowest value of the cosine is -1. The cosine of the angle between two vectors thus determines whether two vectors are pointing in roughly the same direction. In this case, two users correspond to two vectors in the n-dimensional items space. First, the set of items ( $I_{xy}$ ) that both user x and user y have chosen is selected. Then, similarity weights are calculated using the following formula

$$w_{xy} = \frac{\sum_{i \in I_{xy}} r_{xi} r_{yi}}{\sqrt{\sum_{i \in I_{xy}} r_{yi}^2 \sum_{i \in I_{xy}} r_{xi}^2}}, \quad (1)$$

where  $r_{xi}$  is the value of user x on item i and  $r_{yi}$  is the value of user y on item i. In [9] authors utilize a cosine similarity measure between tag vectors to calculate basic similarity of the pages. In [10] authors use cosine similarity to find peer users, and use these peers to recommend resources.

##### B. Pearson correlation

The correlation between two variables reflects the degree to which the variables are related. The most common measure of correlation is the Pearson. Pearson's correlation coefficient reflects the degree of linear relationship between two variables. It ranges from +1 to -1. A correlation

coefficient +1 means that there is a perfect positive linear relationship between variables. In essence, this similarity measure takes into account how much the value of other users for an item deviate from their average value. In this case, similarity between two users  $x$  and  $y$  is measured by computing the Pearson correlation between them using the following formula

$$w_{xy} = \frac{\sum_{i \in I_{xy}} (r_{yi} - r_y)(r_{xi} - r_x)}{\sqrt{\sum_{i \in I_{xy}} (r_{yi} - r_y)^2 \sum_{i \in I_{xy}} (r_{xi} - r_x)^2}}, \quad (2)$$

where  $r_y$  and  $r_x$  denote the average values for users  $x$  and  $y$ , respectively. In essence, this similarity measure takes into account how much the value of other users for an item deviate from their average value. In order to estimate the similarity with the target user, the Pearson correlation coefficient is been proposed to be used [11]. In [12] authors propose to use the Pearson correlation coefficient method for so called „Prominent Items“ recommendation.

### C. Jaccard-Tanimoto index

The Jaccard index is a statistic used for comparing the similarity and diversity of sample sets. The Jaccard measures the overlap degree between two sets by dividing the numbers of items observed by both users (intersection) and the number of different items from both sets of valued items (union). The similarity between two users  $x$  and  $y$  is defined as

$$w_{xy} = \frac{|I_x \cap I_y|}{|I_x| + |I_y| - |I_x \cap I_y|}, \quad (3)$$

where  $|I_x|$  and  $|I_y|$  represent the number of items that have been valued by user  $x$  and user  $y$ , respectively. This similarity metric considers only the number of items that have been valued in common. The metric can be applied on binary datasets that do not contain values. In [13] authors use symmetric measures, like Jaccard, to induce whether two tags have a similar meaning. In [14] authors use Jaccard's coefficient to measure similarity between the new idea and the provided product description.

### D. Sorensen coefficient

The similarity between user  $x$  and  $y$   $\text{sim}(x, y)$  is

$$\text{sim}(x, y) = \frac{2n_{xy}}{n_{x0} + n_{y0}}, \quad (4)$$

where  $n_{xy}$  represents the number of elements both user  $x$  and  $y$  viewed;  $n_{x0}$  is the number of elements that user  $x$  viewed but user  $y$  didn't;  $n_{y0}$  is the number of elements that user  $y$  viewed, but user  $x$  didn't.

## VI. ITEM-BASED COLLABORATIVE FILTERING

The item-based approaches such as [15] apply the same idea, but use similarity items instead of users [16], [17]. When similar items are found, predictions are computed by taking a weighted average of the target user  $v$  values on

these similar. Work [18] results show that item-based techniques hold the promise of allowing CF-based algorithms to scale to large data sets and at the same time produce high-quality recommendations.

## VII. SLOPE ONE SCHEME

The Slope One scheme simplifies the implementation of standard item-based collaborative filtering algorithms and is an alternative to compute predictions. Slope One is a family of algorithms used for collaborative filtering, introduced in [19]. Arguably, it is the simplest form of non-trivial item-based collaborative filtering based on ratings. The simplicity makes it especially easy to implement them efficiently while accuracy is often on par with more complicated and computationally expensive algorithms. Let the set of users who both rated  $x$  and  $y$  be denoted by  $U$ . Given a training set  $c$ , and any two items  $y$  and  $x$  with ratings  $r_{vy}$  and  $r_{vx}$  respectively by some user  $v$  in  $U$ , then the average deviation of item  $x$  with respect to item  $y$  is considered as

$$\text{dev}_{y,x} = \sum_{v \in U} \frac{r_{vy} - r_{vx}}{\text{card}(U)}. \quad (5)$$

The slope one scheme then simplifies the prediction formula to

$$\hat{r}_{vx} = r_v + \frac{1}{\text{card}(R_y)} \sum_{x \in R_y} \text{dev}_{y,x}. \quad (6)$$

The advantage is that this implementation of Slope One does not depend on how the user rated individual items, but only on the user average rating and on which items the user has rated. In [20] authors show, that the accuracy of the algorithm is no longer the only research hotspots. Authors use Slope One in real-time personalized recommendation systems and show, that system characteristics and users' specific needs become two key considerations of these algorithms.

Authors in [21] tried to solve poor quality challenge in collaborative filtering recommender systems and proposed a personalized recommendation algorithm combining slope one scheme and user based collaborative filtering.

## VIII. CONTEXT-AWARE RECOMMENDATIONS

Traditionally recommender systems deal with applications having only two types of entities, users and items, and do not put them into a context when providing recommendations. In [22] authors present heuristic-based approaches and model-based approaches and provide a combined approach of multiple approaches.

In study [23] authors investigate a context-aware recommender system based on rough set theory and collaborating filtering. Based on the assumption that users may have different preference to the recommended items in different contexts, authors proposed their approach to cope with it. The method is compared with an ordinary CF approach and a classical context-aware collaborative filtering approach based on context segments that improve baseline prediction. The experimental results show that their

approach outperforms the others and supplies users a more proactive way that could deliver the proper knowledge to the proper users in the proper context.

## IX. EXPERIMENT

In ViCAMC testing phase we have users with items of e-portfolio. We compute ratings for the items in respect to implicit relevance indications. We split ratings into two sets - observed items and held-out items. Ratings for the held-out items were to be predicted. We used the Mean Absolute Error (MAE) as the evaluation metric for predictive accuracy in this experiment. We compare results to cosine similarity, Pearson correlation, Jaccard index, Sorensen coefficient and slope one results are presented in Fig. 4.

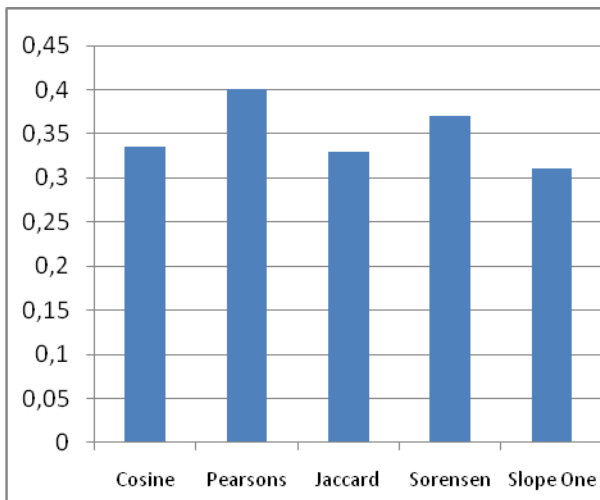


Fig. 4. Use cases for central installation Comparison of correlation results.

## X. CONCLUSIONS

This experiment shows the most relevant method to determine similarity between users and use to implement discussed services. Our analysis confirm the statement, that slope one method accuracy is often on par with more complicated and computationally expensive algorithms. The qualitative estimates of this method – simplicity, accuracy and undemanding calculation expenses motivate to implement it in ViCAMC project.

Appropriate and effective methods for recommendations services to implement must be selected. Analyzed collaborative filtering methods to compute similarity can be implemented in distributed learning environment for collective intelligence utilization. Effective information exchange and collaboration system based on semantic relation discovery method and social networking in R&D community fosters scientific progress and makes economic effect.

## REFERENCES

- [1] A. L. Harris, A. Rea, "Web 2.0 and Virtual World Technologies : A Growing Impact on IS Education", *Journal of Information Systems*, vol. 20, no. 2, pp. 137-145, 2009.
- [2] *Drupal. Web content management system*. [Online]. Available: <http://drupal.org>
- [3] *Conference organization distribution*. [Online]. Available: <http://usecod.com/>
- [4] E. Adomavicius, A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions", *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, 2005. [Online]. Available: <http://dx.doi.org/10.1109/TKDE.2005.99>
- [5] J. L. Herlocker, J. A. Konstan, J. T. Riedl, L. G. Terveen, "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 5-53, 2004. [Online]. Available: <http://dx.doi.org/10.1145/963770.963772>
- [6] G. Lekakos, G. M. Giaglis, "Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors", *Interacting with Computers*, vol. 18, pp. 410-431, 2006. [Online]. Available: <http://dx.doi.org/10.1016/j.intcom.2005.11.004>
- [7] Y. Li, L. Lu, L. Xuefeng, "A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce", *Expert Systems with Applications*, vol. 28, pp. 67-77, 2005. [Online]. Available: <http://dx.doi.org/10.1016/j.eswa.2004.08.013>
- [8] S. L. Wang, C. Y. Wu, "Application of context-aware and personalized recommendation to implement an adaptive ubiquitous learning system", *Expert Systems with Applications*, vol. 38, no. 9, pp. 10831-10838, Sep. 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.eswa.2011.02.083>
- [9] F. Durao, P. Dolog, *A personalized tag-based recommendation in social web systems, Adaptation and Personalization for Web 2.0*, 2009, p. 40.
- [10] J. Gemmell, T. Schimoler, B. Mobasher, R. Burke, "Tag-based resource recommendation in social annotation applications", *User Modeling, Adaptation and Personalization*, Girona, Spain, 2011.
- [11] H. N. Kim, A. T. Jia, I. A. Haa, G. S. Joa, "Collaborative Filtering based on Collaborative Tagging for Enhancing the Quality Recommendation", *Journal of Electronic Commerce Applications*, vol. 9, no. 1, pp. 73-83, 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.elerap.2009.08.004>
- [12] R. Jeet Singh Sodhi, V. Gaur, V. K. Panchal, "Recommendation based on Prominent Items", in *Proc. of the International Conference on Image Information Processing*, no. Icip, 2011, pp. 1-4.
- [13] J. Wang, A. P. de Vries, M. J. T. Reinders, "A User-Item Relevance Model for Log-based Collaborative Filtering", in *Proc. of the European Conference on Information Retrieval*, 2006, pp. 37-48.
- [14] D. Thorleuchter, D. Van den Poel, "Companies Website Optimising concerning Consumer's searching for new Products", in *Proc. of the URKE 2011*, IEEE Press, 2011.
- [15] B. Sigurbjornsson, R. V. Zwol Flickr, "Tag Recommendation based on Collective Knowledge", in *Proc. of the ACM International World Wide Web Conference*, 2008.
- [16] K. Y. Jung, J. H. Lee, "User Preference Mining through Hybrid Collaborative Filtering and Content-based Filtering in Recommendation System", *IEICE Trans.*, vol. E87-D, no. 12, pp. 2781-2790, 2004.
- [17] U. Shardanand, P. Maes, "Social information filtering: algorithms for automating 'word of mouth'", in *Proc. of the Conference on Human Factors in Computing Systems*, 1995, pp. 210-217.
- [18] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms", in *Proc. of the 10th International WWW Conference*, 2001.
- [19] D. Lemire, A. Maclachlan, "Slope one predictors for online rating-based collaborative filtering", in *Proc. of the SIAM Data Mining Conference (SDM '05)*, 2005.
- [20] Z. Sun, N. Luo, W. Kuang, "One real-time personalized recommendation systems based on Slope One algorithm", in *Proc. of the 8th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, Jul. 2011, pp. 1826-1830. [Online]. Available: <http://dx.doi.org/10.1109/FSKD.2011.6019830>
- [21] P. Wang, H. Ye, "A Personalized Recommendation Algorithm Combining Slope One Scheme and User Based Collaborative Filtering", in *Proc. of the International Conference on Industrial and Information Systems*, Apr. 2009, pp. 152-154. [Online]. Available: <http://dx.doi.org/10.1109/IIS.2009.71>
- [22] G. Adomavicius, A. Tuzhilin, "Context-aware recommender systems", in *Proc. of the 2008 ACM conference on Recommender systems - RecSys 08*, 2008, p. 335.
- [23] Z. Huang, X. Lu, H. Duan, "Context-aware recommendation using rough set model and collaborative filtering", *Artificial Intelligence Review*, vol. 35, no. 1, pp. 85-99, Nov. 2010. [Online]. Available: <http://dx.doi.org/10.1007/s10462-010-9185-7>