Revista de Sistemas de Informação da FSMA
n. 12 (2013) pp. 33-40

SISTEMAS DE INFORMAÇÃO

http://www.fsma.edu.br/si/sistemas.html

TRILHA ESTUDANTIL

A Study about Personalized Content Recommendation

Jaqueline Ferreira de Brito, *Graduate student in Information Systems, USP*,
Luciano Antonio Digiampietri, *Professor in the Information Systems Graduate Program, USP*

A Study about Personalized Content Recommendation

***Abstract*—Recommender systems have been proposed as efficient strategies to deal with the information overload which is an Internet reality. In order to study personalized content recommendation, this paper aims at identifying and analyzing, systematically, methods and techniques adopted in the recommendation context. In the systematic review process, 22 papers have been accepted from the main digital libraries in the area. The final results suggest a tendency in the use of strategies to reduce the limitations and to increase the efficiency of the recommendation methods.**

***Index Terms*—Content Recommendation, Personalized Recommendation, Recommender Systems**

I. Introduction

RECOMMENDER systems are software tools and methods that offer suggestions of items that might be useful for someone. The suggestions support the users in their decision processes such as choice of items to buy, what music to hear or which news to read [1, 2]. The following recommendation techniques have received more attention in general: **Content-Based Filtering (CBF):** Recommends to the user items whose content is similar to a content the user read or selected recently; **Collaborative Filtering (CF):** Creates recommendations based on the item evaluations performed by a group of users (neighbors) whose profiles are more similar to the target user; **Hybrid (H):** Combines both of the above mentioned methods.

Besides, the need for personalized content grows every day, for the search for relevant information and minimization of the time spent became an indispensable factor in information search and filtering mechanisms. [1]. Hence, these systems allow for the content to have a well-defined semantic and for the user to better interact with the supplied information given her/his preferences.

This article aims to study the state of the art of content recommendation systems based on user profile. The rest of the article is organized as follows: Section II presents the used methodology, Section III presents the analyzed papers and in Sections IV and V we can see respectively the discussions and the conclusions of this work.

Corresponding author: Luciano Digiampietri, *digiampietri@usp.br*

II. Methodology

IN order to study the current scenario on content recommendation based on user profile, first we performed an exploratory analysis on the subject in order to familiarize ourselves with the main concepts on this topic. Afterward, the systemic review process [3] was made according to the following steps:

1) Planning: step in which we define the research problem and the revision protocol. The current systemic review intends to answer the following question: "Which are the existing techniques for recommendation systems based on the end user profile?".
   For each selected digital library, we considered the following search strings:
   - *ACM Digital*: *Abstract:("content-based" AND "recommender system" AND (profile OR personalize))*
   - *IEEExplore*: *"Abstract":"content-based" AND "recommender system" AND (profile OR personalize)*

   The study used the following selection criteria:
   - We included papers published and available at the scientific digital libraries of *ACM* and *IEEE*;
   - We included recent papers (published in the last five years) that were already approved by the scientific community;
   - We excluded works that present recommendation systems but do not describe the method or technique used;
   - We excluded papers that are not in the knowledge domain of recommendation systems based on end user profile.
2) Conduction: In this step, 39 papers satisfied the inclusion criteria established: 19 from the ACM base and 20 from the IEEE base. Out of those, some were duplicated and others were rejected according to the exclusion criteria defined above [4]. Figure 1, illustrates the status of the papers.
3) Extraction: Step in which 22 papers (5 articles from the ACM base and 17 from the IEEE base) were read and their main characteristics extracted. Table I presents a synthesis of the articles with the established extraction fields.

TABELA I
Synthesis of the selected papers. Source: the authors.

Ref.	Method	Domain	Model	Validation
[5]	H	advertisement	k-NN, k-means, hybrid k-means, ITCC, hybrid ITCC, CCAM, hybrid CCAM	MAE
[6]	CBF	advertisement	k-NN, CBF	user feedback, Precision
[7]	CBF	scientific articles	k-NN, CBF with memory-based and model-based attributes	user feedback
[8]	CBF	e-learning	CBF	TF-IDF nnm b bh, measurement
[9]	H	movies	Naïve Bayes, CF	MAE, F-measure, Coverage
[10]	CBF	movies	Naïve Bayes, semantic CBF	5-fold cross validation
[11]	H	movies	ontology CBF, Apriori	-
[12]	CF	movies	K-medoid, cluster updating algorithm	MAE
[13]	H	movies, TV shows	k-NN, item-based CF, user-based CF, social CBF	students
[14]	CBF	books	trust-based recommendation, Temporal Difference learning	CDC, CDO, CDLOC, CBC - aspect oriented and non-oriented
[15]	CBF	multimedia	smart multimedia	-
[16]	CBF, CF	art museum	Naïve Bayes, Decision Tree, k-NN, RNA, CBF with context, CF with context	Precision, Recall, F-Measure, Coverage, 10-fold cross validation, acceptance test
[17]	H	art museum	Naïve Bayes, genetic algorithm, serendipity heuristic	users
[18]	H	news	Bayesian Framework, CBF, CF	CTR, daily visit frequency
[19]	H	news	semantic CBF, CF	students, TF-IDF measurement
[20]	H	news	Bayesian networks	employees and students
[21]	CBF	researchers	OKAPI, KLD, PM, REL	Average recall, MRR
[22]	CBF	painting	Naïve Bayes, semantic CBF	5-fold cross validation, Precision, Recall
[23]	H	products in general	SOM, item-based CF	RFM measurement
[24]	H	clothes	k-NN, item-based CF, CBF, user-based CF	questionnaire and site information
[25]	H	social bookmarking	WebDCC algorithm, ontology CBF, Rocchio	Precision, Recall
[26]	H	web services	least square algorithm, fuzzy	ANFIS, user feedback

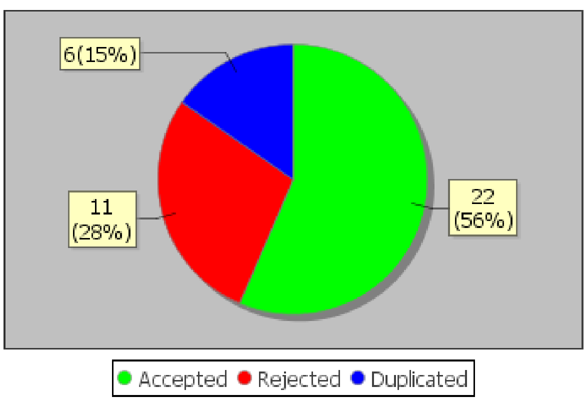

Fig. 1. Status of the selected articles. Source: the authors.

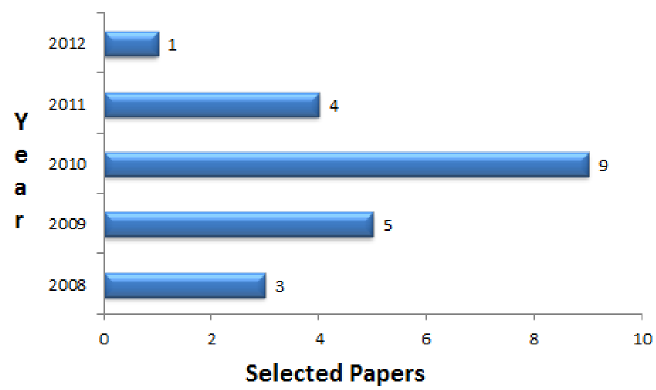

Fig. 2. Distribution of the selected articles through the period of 2008-2012. Source: the authors.

III. Results

Out of the 22 selected articles, most (47,4%) was published in 2010 and a minority (5,3%) in the year of 2012. Figure 2 illustrates the distribution of the selected articles through the last five years.

Besides the user profile and the characteristics of the items, these papers consider other pieces of information in order to formulate the content recommendation:

- **Context information:**
  Based on time variables and group preferences, the model developed by Julashokri and others **[23]** searches for products that are similar to the bought ones and computes the level of interest of the target user, recommending the N most similar products. The measurements R (*Recency*) - interval since the last purchase, F (*Frequency*) - number of purchases in a specific time interval and M (*Monetary*) - amount of money spent by the costumer in that specific period, are considered by the system.
  In **[9]**, the problems of Sparsity Rating (few items evaluated by new users) and Without Contextual Information (absence of information on location, time, day and company) are treated. ModernizeMovie is based on multiple critics to create the pseudo evaluation of movies and on multiple dimensions of context to create the user profile. The evaluation of the Naïve Bayes model developed confirms that the system was able to obtain a higher recommendation precision than the pure CBF and CF strategies.
  Alabastro and others **[16]** deal with the CF and CBF

recommendation methods with and without context information using different algorithms. As interactive guide was developed to recommend personalized visits to an art museum. The best model developed was the collaborative using the K-NN algorithm. The system developed by Iaquinta and others [**17**], also applied to the context of museum itinerary, considers the environment and user behavior information to reorganize the suggested items and make serendipitous recommendations which are surprisingly interesting. A probabilistic profile of the users' interests is built based on the item classification obtained.

In [**20**], a Bayesian network model is presented to recommend personalized news in cell phones. The recommendation is based on activities, preferences, actions, location, time, news content and others to compute the user interest. Because it uses hybrid P2P technologies, the system incorporates the information gathered and automatically offers in real time the news that are considered relevant. Experiments have proved that the method is effective.

- ***Folksonomies***:
  Lops and others [**22**] combine User Generated Content (UGC) with a semantic analysis of the content. One way to deal with UGC is using folksonomy, a taxonomy generated by the users that collaboratively categorize the results with keywords freely chosen (tags). The system estimates the probability of the user becoming interested based on the characteristics of the paintings. Adopting tags increased the precision of the recommendation of works of art, especially for non-specialist users, whose profile became more prominently featured.

  Godoy and Amandi[**25**] propose a strategy that tries to use the user profiles using CBF and frequent used tags. Experiments showed that built profiles overcame recommendation methods based on the popularity of the tags, such as MPTU (Most Popular Tags by User) and MPTR (Most Popular Tags by Resource). Nevertheless, more experiments are needed in order to confirm the preliminary results.

  In [**13**], the author present a hybrid system based on clouds of tags created by the users for movies and TV shows and their classifications for the construction of a cloud of tags for the target user. The model showed a potential efficacy of recommendation for students but its validation phase still needs to be completed.

- **Keywords and Ontologies (or taxonomies):**
  In [**8**], the system learns the user profile implicitly, using the keywords and their automatically created relationships. The relationship among keywords can increase the accuracy in the similarity calculation. This method improves the construction of the user profile and increases the efficacy of the learning with CPF for content recommendation.

  Loh [**6**] investigates the use of keywords (automatically extracted from the text) and classes (extracted from a taxonomy or an ontology) to represent the user profiles. The combination of classes and keywords achieve better results than when they are used in isolation, creating a more accurate profile. The performance of the best strategy was below the other known strategies. Nevertheless, it must be taken into consideration the different experimental conditions that might have affected the final result.

  In [**7**], the authors adopted a model based on the ACM taxonomy to recommend scientific articles for the users of the siteCiteSeerX[1]. The system was able to provide correct recommendations, with no need of extensive records of the user history or explicit rankings.

  Cantador and others [**19**] present the system *News@hand*, which uses semantic technology to recommend news. The contents of the news and the user preferences are described in terms of concepts that show up in a set of domain ontologies. In that paper it is not clear whether the annotation process is done manually or using automatic techniques. The TF-IDF measurement (*Term Frequency-Inverse Document Frequency*) [27] sets weights to all the annotations created for each items, reflecting the importance of each one of them.

  Pan and others [**11**] propose the multi-agent system OARS (Ontology-Based Adaptive Personalized Recommender System) which deals with some of the limitations of the recommender systems: new item, new user, super-specialization, static suggestion and classification sparsity.

  Moursi and others [**15**] propose a decentralized multimedia system (audio, visual and audiovisual) that uses Smart Multimedia (SM), a model that besides having the binary representation of the multimedia file, stores knowledge on the semantic of the internal content, along with the ability of making autonomous decisions. The smart agents are capable of gathering the user and item profile and work collaboratively to deal with the problem of recommendation according to the user's preferences.

- **Others:**
  In [**18**], the Bayesian framework developed uses information from user click logs to forecast the news that will interest the users, based on their activities and the news that have become a trend in the activities of a group of users. The model improved the quality of the recommendation and attracted more visitors to the website Google News.

  In [**24**], a system designed to offer clothing suggestion when a VIP client enters in the store with her/his card (that contains RFID technology). The model has shown the system is an effective mechanism to create high quality recommendations for the customers, but has some limitations: needs a huge database and the calculation of the distance among the customer's

[1] http://CiteSeerx.ist.psu.edu/

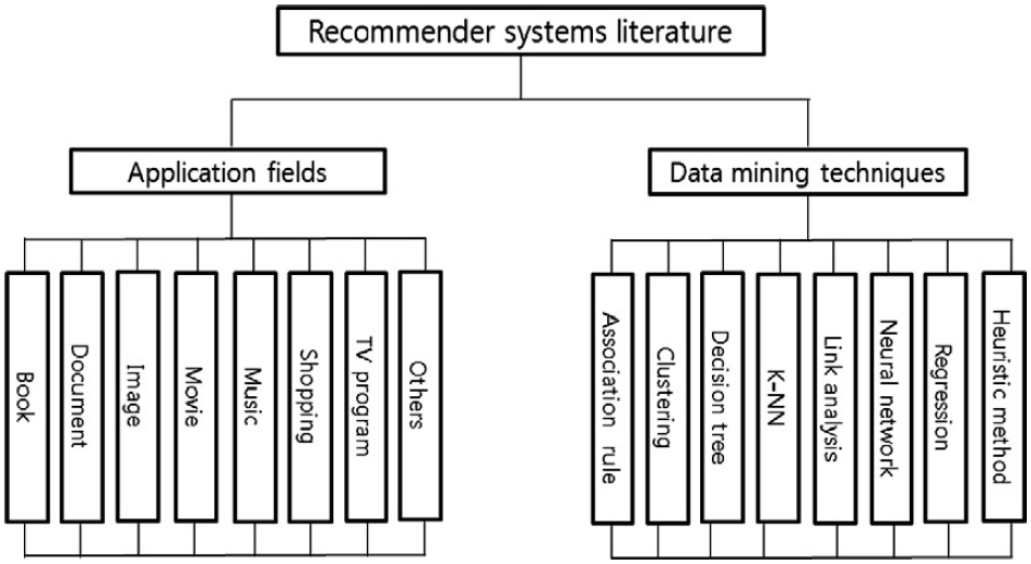

Fig. 3. Structure of the recommendation systems. Source: adapted from Deuk and others [28].

interests takes a long time.

Bedi e Agarwal **[14]** present the AORS (Aspect Oriented Recommender System) that learns the user preferences in a better modular way based on multi-agents. The system validation used the measurements CDC (Concern Diffusion over Components), which indicates the need for more components, CDO (Concern Diffusion over Operations), which indicates the need for more operations (methods or suggestions) for the implementation of the learning based on multi-agents instead of AOP (Aspect Oriented Programming), CD-LOC (Concern Diffusion over LOC), an indicative that the AOP solution is more effective in terms of learning modularization by lines of code, CBC (Coupling Between Component) - for a system with and without aspects. Results point out the learning of the preferences with aspects requires fewer components, less operations to represent the learning and a more effective solution in terms of learning modularization, using the metric of lines of code.

Lops and others **[10]** present a recommendation system for movies in English and Italian, which uses a probabilistic system to classify a new item based on previous preferences. From the experiments performed, the results were very similar to the bilingual and monolingual models, but the model that learn and recommended items in English had a slightly better performance. One of the important considerations is that the error in the disambiguation step introduced in the models must be considered.

In **[26]**, the system is based on non-function properties (response time, reputation, cost, reliability, etc) and on the user preferences to recommend web services. The authors used ANFIS (Adaptive Neuro Fuzzy Inference System) to improve the set of fuzzy rules defined. The system has a unique dynamic characteristic, depending on the feedback of previous users of a specific web service.

Gollapalli and others**[21]** propose a recommendation system for researchers with similar expertise. The explored techniques were: OKAPI, which calculates the similarity between two profiles, KLD (KL Divergence), used to quantify the similarity between two probability distributions, PM (Probabilistic Modeling), which calculates the similarity between two profiles using conditional probability, REL (Trace-based Similarity), which uses a density matrix to compute a relevant score between two profiles. The OKAPI and REL techniques achieve the best results based on the values of average recall and MRR (Mean Reciprocal Rank). Results are indicative of good performance, but still lack exploration of more accurate techniques to represent profiles.

In **[5]**, the co-clustering augmented data matrix model (CCAM) proposed is able to deal with data sparsity. After comparing the performance between models, the authors verifies that the forecast based on models (k-means, ITCC - Information-Theoretic Co-Clustering and CCAM) achieves a much better performance that the forecast based on memory (k-NN). K-Means achieved a much better result in the individual approach than ITCC and CCAM, but the result was closed. Hybrid CCAM achieved a much better result than the Hybrid ITCC and K-Means. This may be justified by the fact that the CCAM has an information gain more useful than the other ones in order to minimize information loss of the multiple related data.

In **[12]**, an incremental clustering approach to improve the scalability of the CF method is presented. Since the clustering reconstruction technique is too expensive, an algorithm was proposed to dynamically update the clusters. Comparing the values of MAE (Mean Absolute Error) found by the dynamic situations and by the situation before the initial clustering, the dynamic approach found a similar result to the initial situation, but showed itself more advantageous when new data entered the system.

IV. Discussion

COMPARING Table I and Figure 3, we can see that application domains and data mining techniques or models converge. The main models applied in the analyzed papers that brought more quality and precision to the results are k-NN, Naïve Bayes and the algorithms CF and CBF, given the different application domains involved.

Even though the performance of the Naïve Bayes method is not as good as some of the other static learning methods, such as the k-NN and SVM (Support Vector Machines) classifiers, we verified that it can achieve a surprisingly good performance in classification techniques in which the update of the computed probability is not important. Other advantages of the Naïve Bayes approach is that it is very efficient and easy to implement when compared to other learning methods [29].

The recommendation method that is more used in the analyzed papers is the hybrid one, because it has the advantages of the CBF and CF methods. In Table II one can see the advantages and limitations of each method [30].

The CBF method does not need the user evaluation on an item. This method compares all available products in the database with the user profile. Nevertheless, the item characteristics that deserve special attention are not taken into consideration, since the user does not provide notes in this type of filter. Hence, all item characteristics are treated equally. The problem of overspecialization is another limitation, because the generated recommendations are similar to the keywords found in the user profile.

The CF method is based on the user evaluations on the items, independent of the nature of the content we intend to recommend. The quality of the recommendation generated by this method depends on the evaluations (whether positive or negative) given by the users. Besides, this method is capable of making serendipitous recommendation. Nevertheless, a product that was never evaluated cannot be recommended and if the database is too big, it is necessary for the user to buy or evaluate a considerable number of products in order to make it possible to find good neighbors with similar performances. Users that evaluate less than 50 items can suffer from the false neighbor problem because they may coincidentally evaluate the same items, but since their profiles have few items, it is impossible to generalize the similarity. Thus, the more products are equally evaluated, the higher the chances the profiles are really similar. In order to create groups with similar tastes, it is necessary to calculate the similarity of a target user with all other users, which may be a process that is long and computationally costly.

The disadvantages of the mentioned methods are complementary, that is, the disadvantages of one are the advantages of the other and vice versa. Hence, the hybrid method is able to put together both methods, usually with the help of other techniques, in order to minimize the limitations of each one of them. The sole disadvantage is the startup problem, which refers to the time the system takes to find relevant information on the user to generate good recommendations. Some authors suggest recommend a list of the most sold items or to use the user's demographic data (age, sex, city, etc.) to generate the initial recommendations.

In this review we found out that the focus of the reviewed papers is more on the representation of the user profile than on the recommendation process itself, given that an improve on the quality of the former will also improve the quality and precision of the latter. The papers suggest that there is a trend to use context information, folksonomies, serendipity property, keywords, ontologies (or taxonomies) among other strategies, in order to diminish the limitations and improve the efficiency of the applied methods. Besides, most reviewed papers present validation strategies for the built models, highlighting the importance of this step of the process when analyzing the built systems.

V. Conclusions

RECOMMENDER systems have called the attention of professionals and academics because they are tools that support the users in several decision making contexts when faced with huge amounts of information available at the Web.

Thus, we identified that the most used method in all papers is the hybrid one, due to the fact that it joins the advantages of the collaborative and content based recommendation methods.

Besides, the most frequently used methods in the learning step were k-NN, Naïve Bayes and the algorithms CF and CBF. Generally, these methods achieved a higher efficacy result among the applied models.

Most papers validated the proposed solutions, whether through experimentation with participating students, sampling techniques or some specific type of statistic measurement. We also observed that the focus of the reviewed papers is more on the representation of the user profile than on the recommendation process itself, reflecting a global quality improvement of the system.

Hence, the papers point out the fact that besides the user profile and the item characteristics, they consider other information in the content recommendation process: context information, folksonomies, serendipitous property, keywords, ontology (or taxonomies) and others. This suggests a trend indicative, given that those are strategies to minimize limitations and, at the same time, improve the efficiency of the learning of the main recommendation methods available.

TABELA II
ADVANTAGES AND LIMITATIONS OF THE RECOMMENDATION METHODS. SOURCE: THE AUTHORS.

Method	Advantage	Limitation
CBF	does not suffer from the first evaluator problem; all items may be recommended	does not consider aspect such as text quality and author renown; overspecialization
CF	independence from the content; generation of recommendation based on user preferences; possibility of generating unexpected high quality recommendations	the problem of the first evaluator; database dispersion; black sheep (false neighbor); processing cost
H	combines the advantages and minimized the individual limitations of both methods above mentioned	the startup problem

References

[1] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds., *Recommender Systems Handbook*. Springer, 2011, ISBN: 978-0-387-85819-7.

[2] S. C. Cazella, M. A. S. N. Nunes, and E. Reategui, "Jornada de atualização de informática," in. SBC, 10, ch. A Ciência do Palpite: Estado da Arte em Sistemas de Recomendação, pp. 161–216.

[3] J. Biolchini, P. G. Mian, and A. C. C. Natali, "Systematic review in software engineering," COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, Tech. Rep. RT-ES 679/05, 2005.

[4] A. Zamboni, A. Thommazo, E. Hernandes, and S. Fabbri, *Start: uma ferramenta computacional de apoio à revisão sistemática*, 2012. [Online]. Available: http://lapes.dc.ufscar.br/software/start-tool.

[5] M.-L. Wu, C. H. Chang, and R. Z. Liu, "Collaborative filtering with ccam," in *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, vol. 2, 2011, pp. 245 –250. DOI: 10.1109/ICMLA.2011.47.

[6] S. Loh, F. Lorenzi, G. Simões, L. K. Wives, and J. P. M. de Oliveira, "Comparing keywords and taxonomies in the representation of users profiles in a content-based recommender system," in *Proceedings of the 2008 ACM symposium on Applied computing*, ser. SAC '08, Fortaleza, Ceara, Brazil: ACM, 2008, pp. 2030–2034, ISBN: 978-1-59593-753-7. DOI: 10.1145/1363686.1364177. [Online]. Available: http://doi.acm.org/10.1145/1363686.1364177.

[7] A. Kodakateri Pudhiyaveetil, S. Gauch, H. Luong, and J. Eno, "Conceptual recommender system for citeseerx," in *Proceedings of the third ACM conference on Recommender systems*, ser. RecSys '09, New York, New York, USA: ACM, 2009, pp. 241–244, ISBN: 978-1-60558-435-5. DOI: 10.1145/1639714.1639758. [Online]. Available: http://doi.acm.org/10.1145/1639714.1639758.

[8] X. Wan, Q. Jamaliding, F. Anma, and T. Okamoto, "Applying keyword map based learner profile to a recommender system for group learning support," in *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on*, vol. 1, 2010, pp. 3 –6. DOI: 10.1109/ETCS.2010.439.

[9] N. Rattanajitbanjong and S. Maneeroj, "Multi criteria pseudo rating and multidimensional user profile for movie recommender system," in *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*, 2009, pp. 596 –601. DOI: 10.1109/ICCSIT.2009.5234808.

[10] P. Lops, C. Musto, F. Narducci, M. de Gemmis, P. Basile, and G. Semeraro, "Learning semantic content-based profiles for cross-language recommendations," in *Proceedings of the First Workshop on Personalised Multilingual Hypertext Retrieval*, ser. PMHR '11, Eindhoven, Netherlands: ACM, 2011, pp. 26–33, ISBN: 978-1-4503-0897-7. DOI: 10.1145/2047403.2047409. [Online]. Available: http://doi.acm.org/10.1145/2047403.2047409.

[11] P.-Y. Pan, C.-H. Wang, G.-J. Horng, and S.-T. Cheng, "The development of an ontology-based adaptive personalized recommender system," in *Electronics and Information Engineering (ICEIE), 2010 International Conference On*, vol. 1, 2010, pp. V1–76 V1–80–. DOI: 10.1109/ICEIE.2010.5559842.

[12] P. Chakraborty, "A scalable collaborative filtering based recommender system using incremental clustering," in *Advance Computing Conference, 2009. IACC 2009. IEEE International*, 2009, pp. 1526 –1529. DOI: 10.1109/IADCC.2009.4809245.

[13] A. Barragans-Martinez, M. Rey-Lopez, E. Costa Montenegro, F. Mikic-Fonte, J. Burguillo, and A. Peleteiro, "Exploiting social tagging in a web 2.0 recommender system," *Internet Computing, IEEE*, vol. 14, no. 6, pp. 23 –30, 2010, ISSN: 1089-7801. DOI: 10.1109/MIC.2010.104.

[14] P. Bedi and S. Agarwal, "Preference learning in aspect-oriented recommender system," in *Computational Intelligence and Communication Networks (CICN), 2011 International Conference on*, 2011, pp. 611–615. DOI: 10.1109/CICN.2011.132.

[15] S. Moursi, M. Elsakhawy, and H. Ghenniwa, "Agent oriented media recommender system utilizing smart multimedia," in *Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference on*, 2011, pp. 437 –444. DOI: 10.1109/CSCWD.2011.5960110.

[16] P. Alabastro, M. Ang, R. de Guzman, M. Muhi, and M. Suarez, "Mymuseum: integrating personalized recommendation and multimedia for enriched human-system interaction," in *Digital Content, Multimedia Technology and its Applications (IDC), 2010 6th International Conference on*, 2010, pp. 421 –426.

[17] L. Iaquinta, M. de Gemmis, P. Lops, and G. Semeraro, "Recommendations toward serendipitous diversions," in *Intelligent Systems Design and Applications, 2009. ISDA '09. Ninth International Conference on*, 2009, pp. 1049 –1054. DOI: 10.1109/ISDA.2009.215.

[18] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior," in *Proceedings of the 15th international conference on Intelligent user interfaces*, ser. IUI '10, Hong Kong, China: ACM, 2010, pp. 31–40, ISBN: 978-1-60558-515-4. DOI: 10.1145/1719970.1719976. [Online]. Available: http://doi.acm.org/10.1145/1719970.1719976.

[19] I. Cantador, A. Bellogin, and P. Castells, "Ontology-based personalised and context-aware recommendations of news items," in *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on*, vol. 1, 2008, pp. 562 –565. DOI: 10.1109/WIIAT.2008.204.

[20] K. F. Yeung and Y. Yang, "A proactive personalized mobile news recommendation system," in *Developments in E-systems Engineering (DESE), 2010*, 2010, pp. 207 –212. DOI: 10.1109/DeSE.2010.40.

[21] S. D. Gollapalli, P. Mitra, and C. L. Giles, "Similar researcher search in academic environments," in *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, ser. JCDL '12, Washington, DC, USA: ACM, 2012, pp. 167–170, ISBN: 978-1-4503-1154-0. DOI: 10.1145/2232817.2232849. [Online]. Available: http://doi.acm.org/10.1145/2232817.2232849.

[22] P. Lops, M. de Gemmis, G. Semeraro, P. Gissi, C. Musto, and F. Narducci, "Content-based filtering with tags: the first system," in *Intelligent Systems Design and Applications, 2009. ISDA '09. Ninth International Conference on*, 2009, pp. 255 –260. DOI: 10.1109/ISDA.2009.84.

[23] M. Julashokri, M. Fathian, and M. Gholamian, "Improving customer's profile in recommender systems using time context and group preferences," in *Computer Sciences and Convergence Information Technology (ICCIT), 2010 5th International Conference on*, 2010, pp. 125 –129. DOI: 10.1109/ICCIT.2010.5711042.

[24] Y. Liangxing and D. Aihua, "Hybrid product recommender system for apparel retailing customers," in *Information Engineering (ICIE), 2010 WASE International Conference on*, vol. 1, 2010, pp. 356 –360. DOI: 10.1109/ICIE.2010.91.

[25] D. Godoy and A. Amandi, "Hybrid content and tag-based profiles for recommendation in collaborative tagging systems," in *Latin American Web Conference, 2008. LA-WEB '08.*, 2008, pp. 58 –65. DOI: 10.1109/LA-WEB.2008.15.

[26] S. Tiwari and S. Kaushik, "A non functional properties based web service recommender system," in *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on*, 2010, pp. 1 –4. DOI: 10.1109/CISE.2010.5676953.

[27] D. W. Oard, "The state of the art in text filtering," *User Modeling and User-Adapted Interaction*, vol. 7, no. 3, pp. 141–178, Mar. 1997, ISSN: 0924-1868. DOI: 10.1023/A:1008287121180. [Online]. Available: http://dx.doi.org/10.1023/A:1008287121180.

[28] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, "A literature review and classification of recommender systems research," *Expert Systems with Applications*, vol. 39, no. 11, pp. 10059 –10072, 2012, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2012.02.038. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417412002825.

[29] P. Lops, M. de Gemmis, and G. Semeraro, "Content-based recommender systems: state of the art and trends.," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds., Springer, 2011, pp. 73–105, ISBN: 978-0-387-85819-7. [Online]. Available: http://dblp.uni-trier.de/db/reference/rsh/rsh2011.html#LopsGS11.

[30] V. M. Filho, "E-recommender: sistema inteligente de recomendação para comércio eletrônico," *Trabalho de Conclusão do Curso de Engenharia da Computação*, 2006.

Jaqueline Ferreira de Brito is graduated on Information Systems by the School of Arts, Sciences and Humanities from the University of São Paulo, EACH-USP (2010). In 2012 she became a graduate students at the masters program on Information Systems at EACH-USP. She has experience in Artificial Intelligence, with emphasis of Artificial Neural networks and Pattern Recognition.

Luciano Antonio Digiampietri (corresponding author) is graduated in Computer Science from the State University of Campinas (2002) and a PhD in Computer Science from the same institution (2007). Since April 2008 he is a researcher professor in the Bachelor Program on Information Systems at the School of Arts, Sciences and Humanities from the University of São Paulo (EACH-USP) and since 2010 he is a permanent professor at the Information Systems Master program at the University of São Paulo. He has experience in the area of Computer Science, with emphasis on Computational Biology, Databases and Artificial Intelligence, working specially in the following subjects: scientific workflows, bioinformatics, data provenance, automatic service composition, experiment tracking and algorithms .
E-mail: *digiampietri@usp.br*.