

# Generating Context-aware Recommendations using Banking Data in a Mobile Recommender System

Daniel Gallego Vico, Gabriel Huecas and Joaquín Salvachúa Rodríguez

*Departamento de Ingeniería de Sistemas Telemáticos*

*Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid*

*Avenida Complutense 30, 28040, Madrid, Spain*

*Email: {dgallego, gabriel, jsalvachua}@dit.upm.es*

**Abstract**—The increasing adoption of smartphones by the society has created a new area of research in recommender systems. This new domain is based on using location and context-awareness to provide personalization. This paper describes a model to generate context-aware recommendations for mobile recommender systems using banking data in order to recommend places where the bank customers have previously spent their money. In this work we have used real data provided by a well know Spanish bank. The mobile prototype deployed in the bank Labs environment was evaluated in a survey among 100 users with good results regarding usefulness and effectiveness. The results also showed that test users had a high confidence in a recommender system based on real banking data.

**Keywords**-Mobile Recommender; Context-aware; Banking data mining; User modeling; Customer segmentation

## I. INTRODUCTION

In recent years the mobile world has evolved extremely quickly not only in terms of adoption, but also in technology. The result of these advances is a high adaptive personalization of mobile applications. These new capacities provided by smartphones give rise to the possibility of building enhanced mobile commerce applications using all the user data we have at our disposal by utilizing their context sensors.

On the other hand, the eBusiness world has also advanced due to this new way of personalization. Good examples of this evolution are recommender systems. Traditional recommender systems usually are based on subjective data or personal scores provides by the users (e.g. Google Places). However, in recent years new platforms have based their recommendation on real purchases and therefore, the recommendations inspire more confidence (e.g. Amazon). This confidence in the results is always a key feature in any recommender system, but usually it is not easy to have such kind of data from real purchases. As a result, if we think in bank entities, we will probably agree that they are one of the best sources of trusted data in the world, as they have a huge amount of transactions from millions of users.

In this paper we present a mobile prototype based on using banking data to generate enhanced context-aware recommendations. This research project was carried out through the

collaboration between our research group and one of the most important Spanish banks (its identity is not revealed in these lines in order to comply with bank's policies). This banking entity has provided us with more than 2.5 million credit card transactions made during the year 2010 and information about the 222,000 places and 34,000 anonymous customers' profiles related to the previous transactions.

The rest of the paper is organized as follows: the next section reviews related work. Section 3 describes the motivations behind this research. Section 4 presents the model used to generate the context-aware recommendations. Section 5 provides the results of our experimental work based on the prototype deployment and the survey carried out. After that, in Section 6 we discuss the results achieved. Finally, we conclude with a short summary and directions for future research.

## II. RELATED WORK

A large amount of research and practical applications exist on mobile computing, recommender systems, context-awareness (e.g. [1] or [2]) or location based services, as well as any combination of the above areas. For instance, Kenteris et al. recently surveyed the field of mobile guides [3]. Ricci also discusses the goals of context-dependent recommendations and their importance in mobile recommender systems in his recent survey [4].

However, as Yujie and Licai stressed in [5] one of the most important challenges for context-aware recommender systems is the lack of public datasets available to conduct experimental evaluations on the methods developed.

On the other hand, it is important to note that usually all of the projects related to a banking data mining process in a bank entity are focused on generating mined knowledge useful for the bank workers, helping them to make decisions about customer segmentation or economic products, as we can see in [6] and [7]. Therefore, we cannot find research work in which the banking data is used to generate recommendations to the end users.

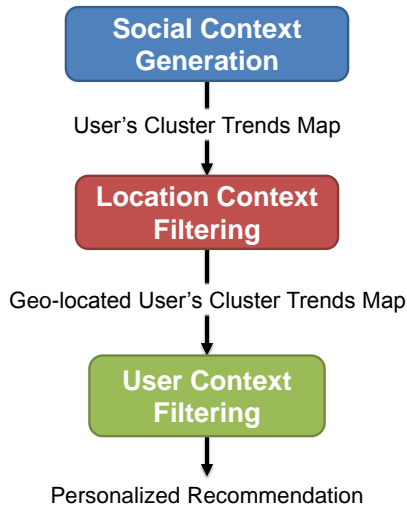


Figure 1. Adaptive recommendation process.

### III. MOTIVATION

All the important banks have millions of records in their databases plenty of rich information about customer purchases, client profiles or economic trends. Nevertheless, the vast majority of them frequently do not use correctly or underutilize these data to achieve a maximum benefit for their clients. Sometimes this is so because the privacy and security policies related to these data are complex to manage. In other cases, the challenge is related to a data mining scalability problem due to the huge data available to be processed.

Bearing this in mind, the emerging of recommender systems in this environment in response to these problems is a direct consequence. We have developed a method to generate context-aware recommendations for mobile recommender systems based on banking data. With this enhanced context-awareness our aim is to recommend *places*. A place is any entity where bank clients have paid with their credit cards (e.g. restaurants, stores, cinemas, supermarkets and so on).

Consequently, we have achieved a novelty application in the banking environment. This extra value provided to the end users is essential for our Spanish banking partner as it is an advantage in terms of market competition. It is also important to point out that using these banking data span across a wide domain range of recommendations categories, whereas most prior work tends to be more narrowly, often focused on a single store or a small set of products. Hence opportunities arise for cross-domain recommendations due to the richer context is possible to generate using banking data.

An additional main idea behind this research is the confidence on the recommendations generated. As Swearingen and Sinha [8] and Tintarev and Masthoff [9] said, one of the key goals of every recommender system is achieve the

*trust* property to increase the users' confidence in the system recommendations. When we usually use a recommender system, we can think about several ways of falsifying or distorting the reality related to the items recommended. For example, the score of a restaurant recommendation from Google Places [10] is based on different user opinions. Thus the final recommendation is based on subjective evaluations of each user, and in some cases, the recommendation might not correspond to the reality. In conclusion, sometimes you might not trust recommendations because of the doubtful data origin. In our case we accomplish this goal because the system inspires confidence, as the data used for recommending are real data from the bank.

### IV. CONTEXT-AWARENESS GENERATION

As we mentioned in Section 2, there has been much research on the area of generating context-awareness and different definitions of the term *context* exist (e.g. [11], [12]). Therefore, we follow the definition proposed by Dey [13]: "*Context is any information that can be used to characterize the situation of an entity*". Specifically, the context dimensions in which our system is based on are: Social, Location and User context.

In the following sections we are going to present how we generate and use them to improve the recommendations, describing in detail the adaptive recommendation process summarized by Figure 1.

#### A. Social Context

The social context is generated by a data mining process over the banking data divided into three steps (Figure 2). These steps are not constricted to a real-time execution because all of them are carried out before the recommendations are requested by the user.

In the first step (User Profile Clustering), the system takes the banking client profiles provided to apply a clustering segmentation to them. These data have to be cleaned before the processing starts, so each record containing incorrect data (e.g. incorrect format, missing values, etc.) are ignored in the clustering process. It is very important to point out that only a restricted set of information was provided by our Spanish banking partner from its databases, being also previously anonymized in order to avoid privacy problems and to comply with the bank's policies. For this reason, the client profiles provided by the bank entity are represented by the following straightforward  $n$ -tuple:

$$\langle profileID, gender, age, averageExpensePerYear \rangle \quad (1)$$

In order to reduce the complexity of the banking data mining process, we first apply a Canopy clustering process [14] on these data and then a K-means clustering process [15] over the canopies generated, achieving a set of clusters based on the similarity of the banking clients. We have

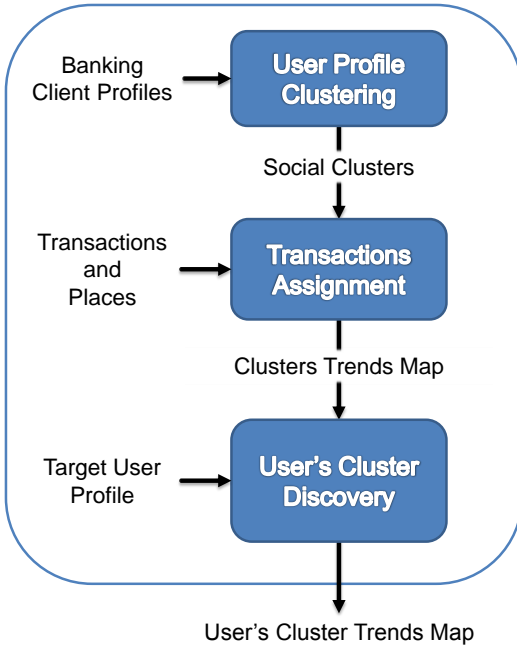


Figure 2. Social context generation process.

called this set of clusters *Social Clusters*, because they gather together banking clients with similar profiles, forming social groups where the consumption model or tastes are related.

In the Transactions Assignment step, the system first assigns the credit card transactions to the corresponding cluster, considering that there is a unequivocal relationship between a transaction and a client (given by the *profileID* element that indicated who made that credit card transaction). Every bank transaction is represented by the following *n*-tuple:

$$\langle profileID, placeID, paymentAmount, time, date \rangle \quad (2)$$

After that, all the transactions are assigned to the social clusters and then, a second process identifies the places where the transactions were made. The places are represented by the following *n*-tuple:

$$\langle placeID, category, name, address, latitude, longitude \rangle \quad (3)$$

With this second process, we create a map of places where the relationships among places and clusters are shown, noticing in this way the consumption trends of every cluster.

Finally, the User's Cluster Discovery process is activated when the user enters the first time to the mobile application. The system checks the information profile extracted from the user's banking account (a *n*-tuple like the one show in 1) in order to assign her to any of the existing social clusters. This is carried out by calculating the distance among the point

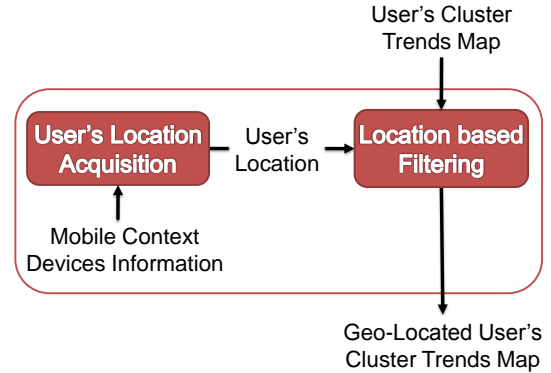


Figure 3. Location context filtering process.

that represents the user profile in the space defined and the centroids from every social cluster. A centroid is a virtual point corresponding to the average of all the real points in the cluster. That is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. Hence, the cluster with the centroid at a minor distance from the user profile point representation is the social cluster assigned to the user.

After these steps we know the social context of the user because now she has been assigned to one of the social clusters generated. Every cluster has a common consumption model represented by the Clusters Trends Map and thus, we know which places are candidates to be recommended to her.

For instance, if a user belongs to the social cluster of 50 year-old women with an average expenditure of EUR 10,000 per year in credit card purchases, the set of possible places to recommend is made up of the places in which people in this category have paid with their credit cards in the year 2010 (as the data provided by our bank partner for this research correspond to that year).

### B. Location Context

As [11] said, location is currently one of the most important context parameters. Accordingly, after obtaining the social context of the user based on the banking information, the recommendation can be made more accurate by adding the location context dimension. Most of the time, end users are looking for places recommendations in their immediate locality (e.g. good restaurants nearby). The use of mobile context device information as an input for the recommender system allows us to personalize recommendations based on the user's location.

Different mobile context devices are involved in the acquisition of the user's location. If the user's device is GPS-capable, the geo-location will be more accurate. If not, a less accurate but usually valid location can be obtained using network-based positioning technologies or Geo IP capabilities.

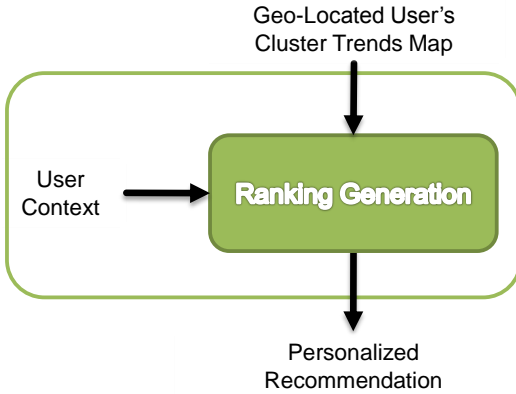


Figure 4. User context filtering process.

Once the system is aware of the user's location, it is applied as a new input to filter the user's cluster trends map, obtaining a geo-located user's cluster trends map (Figure 3).

### C. User Context

The final process to achieve the personalized recommendation takes into account the user context. This context is based on a set of parameters (e.g. current time or current activity of the user inferred from sensor data) and an input preference given by the user to know the place category (one of the elements of the  $n$ -tuple 3) she wants to be recommended (e.g. restaurant, supermarket, cinema, etc.).

For instance, if the user wants a restaurant recommendation (category input), the mobile application could also use the current time information (e.g. lunch time) to filter the geo-located user's cluster trends map (Figure 4) considering only those restaurants that fit with her current activity (e.g. walking). Following with the example, the user would see only a ranking of the closest restaurants at walking distance to her location that the banking clients belonging to her social cluster has visited the most at lunch time. That ranking would be generated by ordering those restaurants attending to the number of customers that have previously visited everyone.

## V. EVALUATION AND RESULTS

To evaluate the system, a prototype was developed and deployed in a real environment that belongs to our bank partner called *Labs*. The primary aim of *Labs* is to allow the deployment of new researches and development projects created in the bank in order to be able to collect feedback from a set of bank clients registered in this environment.

Using this platform we evaluated first the social clusters achieved after applying the previous processes to real banking data. Then, we set up an online survey with two scenarios using a mobile prototype developed in Android in order to evaluate the user acceptance.

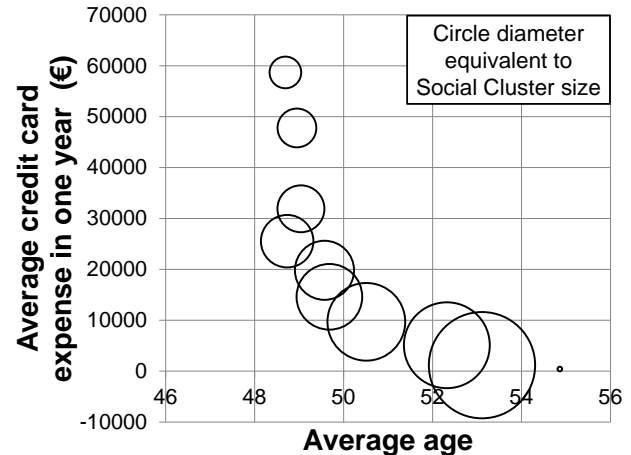


Figure 5. Social Clusters distribution by average expense, age and size.

### A. Social Clusters

The banking data provided by our bank partner to create the social clusters consisted of more than 2.5 million credit card transactions made during the previous year, providing information on 222,000 places and 34,000 anonymous customers' profiles from customers between 48 and 55 years old. All these data were provided by the bank following the  $n$ -tuples (1), (2) and (3).

The Figure 5 illustrates the social clusters emerged after the clustering process over the banking data. As we can see, the average credit card expense in one year, the average age and the size of every social cluster (given by the diameter of the circles) is shown.

### B. User Acceptance

We have analyzed the feedback provided by 100 bank customers registered in the *Labs* platform where the system is deployed. This evaluation was carried out using an on line questionnaire based on two scenarios. The first one was focused on restaurant recommendations and the second one on supermarket recommendations. Both scenarios show a simple case in which after a user request, she receives a recommendation compose by several places corresponding to the previous categories. Figure 6 depicts a screenshot for a lunch recommendation provided by the Android mobile prototype taking into account that the location is provided by the mobile context sensors and the user context information is previously provided by the user.

Therefore, after a brief experience with the application in those scenarios the test users were asked to judge several statements related to some properties using a 5-point scale, where 1 mean "totally disagree" and 5 "totally agree". The statements were like this: "The application is [property evaluated]". Additionally, users had free text inputs fields to



Figure 6. Mobile application interface for a lunch recommendation.

make comments and annotations. The results are illustrated with average values in Figure 7.

## VI. DISCUSSION

First of all, if we think about the distribution of customers along the social clusters (Figure 5), the results confirm the intuition: the clusters with less people are the clusters who spent more money in credit card transactions because high economic class people are more infrequent than medium and low economic class people. While the bigger size clusters are those who spend less money and also, are composed by older people that are less used to pay with credit card than younger people.

In regard to the results related to the user acceptance, they reveal a very positive attitude towards the mobile recommender system shown in the Figure 6, as long as it has average high scores in all the properties analyzed.

On the other hand and attending to the way we manage the explanations in our recommender system, it achieves some of the most important criteria recently set out by Tintarev and Masthoff in [9]. Specifically the “transparency” (i.e. explain how the system works) is achieved due to the explanation provided for the places recommended, as the application informs the user about how the recommendations have been generated considering the purchases of other bank customers like her. The “trust” (i.e. increase users’ confidence in the system), “effectiveness” (i.e. help users

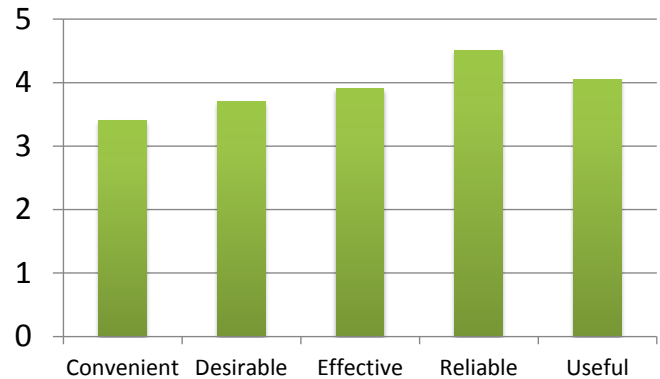


Figure 7. User acceptance survey results.

make good decisions) and “satisfaction” (i.e. increase the ease of use or enjoyment) criteria are achieved if we take into account the high values of the “reliable”, “effective” and “useful” properties respectively evaluated in the survey (Figure 7).

The statistical outcome is also supported by the comments wrote by the test users during the survey process. For example: “*Having a recommender system in my smartphone available anywhere, anytime for searching any kind of place is very usefull.*”, “*I think that the most important feature for me is the confidence in the results as they come from real data of my own bank*” or “*I will really appreciate to have such a kind of application in my mobile phone in my daily life.*”.

However, some of them remarked the privacy issues related to deploy this system into a real commercial exploitation, as they did not want their personal data in danger. Although the system is right now deployed in the bank Labs environment (a close secure environment), this point is an important issue that has to be studied if the system is deployed in a future outside of it, attending to the facts pointed out by Ohm in [16], where he showed that in some cases anonymization is not enough to preserve privacy promises.

## VII. CONCLUSION AND FUTURE WORK

In this paper, we have presented a method of generating context-aware recommendations using banking data in mobile recommender systems. As we have shown, using this kind of data based on real people actions and banking history, allows us to increase the confidence in the personalized recommendations generated, because there is no subjective data used in the recommendation process. This feature of our system provides an essential advantage compared to other recommender systems which have the aforementioned problem of being based on non-reliable data.

Current and future work includes the evolution of the current prototype into a complete mobile application. This would allow us to evaluate the system with users really

interacting with a mobile device in realistic scenarios related to the users' daily life. Of course, the long run aim of our banking partner is to launch a real product for commercial exploitation based on our system. Thus it will be necessary to study the privacy problems related to that real deployment.

On the other hand, we want to analyze the impact of using proactive techniques in our recommendation process. Proactivity means that the system pushes recommendations to the user when the current situation seems appropriate without being needed a explicit user request. Therefore, we are working now in a model to achieve proactivity in mobile context-aware recommender systems ([17] and [18]) that could be integrated into the context-generation model presented in this paper.

Another open issue that could be studied in relation with enhancing the recommendation process is the one described in [19], in order to create multiple personalities in the system that would have their own personalized profile, like a "what kind of customer do you want to be today?" feature. As a result, a user of the system could have several profiles with different social clusters associated. This is an interesting feature if we bear in mind that sometimes people pay with their credits cards to buy gifts or services for friends or family that usually does not have the same tastes.

#### ACKNOWLEDGMENT

The authors would like to thank the bank *Labs* group for providing their banking data and their expertise on the banking domain without which this research could not have been possible.

#### REFERENCES

- [1] M. Baldauf, S. Dustdar, and F. Rosenberg, "A survey on context-aware systems," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 2, pp. 263–277, June 2007. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1356236.1356243>
- [2] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in *Proceedings of the 2008 ACM conference on Recommender systems*, ser. RecSys '08. New York, NY, USA: ACM, 2008, pp. 335–336. [Online]. Available: <http://doi.acm.org/10.1145/1454008.1454068>
- [3] M. Kenteris, D. Gavalas, and D. Economou, "Electronic mobile guides: a survey," *Personal Ubiquitous Comput.*, vol. 15, pp. 97–111, January 2011. [Online]. Available: <http://dx.doi.org/10.1007/s00779-010-0295-7>
- [4] F. Ricci, "Mobile recommender systems," *International Journal of Information Technology and Tourism*, 2011.
- [5] Z. Yujie and W. Licai, "Some challenges for context-aware recommender systems," in *Computer Science and Education (ICCSE), 2010 5th International Conference on*, aug. 2010, pp. 362–365.
- [6] P. Ataei, "Mining the (data) bank," *Potentials, IEEE*, vol. 24, no. 4, pp. 40–42, 2005.
- [7] S. Ren, Q. Sun, and Y. Shi, "Customer segmentation of bank based on data warehouse and data mining," in *Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*, 2010, pp. 349–353.
- [8] K. Swearingen and R. Sinha, "Interaction design for recommender systems," in *In Designing Interactive Systems 2002*. ACM Press, 2002.
- [9] N. Tintarev and J. Masthoff, "Designing and evaluating explanations for recommender systems," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer US, 2011, pp. 479–510.
- [10] Google, "Places," 2011. [Online]. Available: <http://www.google.com/hotpot>
- [11] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a better understanding of context and context-awareness," in *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*, ser. HUC '99. London, UK: Springer-Verlag, 1999, pp. 304–307. [Online]. Available: <http://portal.acm.org/citation.cfm?id=647985.743843>
- [12] M. Bazire and P. Brézillon, "Understanding context before using it," in *CONTEXT'05*, 2005, pp. 29–40.
- [13] A. K. Dey, "Understanding and using context," *Personal Ubiquitous Comput.*, vol. 5, pp. 4–7, January 2001. [Online]. Available: <http://dx.doi.org/10.1007/s007790170019>
- [14] A. McCallum, K. Nigam, and L. H. Ungar, "Efficient clustering of high-dimensional data sets with application to reference matching," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '00. New York, NY, USA: ACM, 2000, pp. 169–178. [Online]. Available: <http://doi.acm.org/10.1145/347090.347123>
- [15] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 881–892, Jul. 2002.
- [16] P. Ohm, "Broken promises of privacy: Responding to the surprising failure of anonymization," *Social Science Research Network*, vol. 57, pp. 1–64, 2009.
- [17] W. Woerndl, J. Huebner, R. Bader, and D. Gallego-Vico, "A model for proactivity in mobile, context-aware recommender systems," in *the 5th ACM International Conference on Recommender Systems*, October 2011.
- [18] D. Gallego-Vico, W. Woerndl, and R. Bader, "A study on proactive delivery of restaurant recommendations for android smartphones," in *the International Workshop on Personalization in Mobile Applications (PeMA) at 5th ACM International Conference on Recommender Systems*, October 2011.
- [19] L. F. Cranor, "'i didn't buy it for myself' privacy and ecommerce personalization," in *Proceedings of the 2003 ACM workshop on Privacy in the electronic society*, ser. WPES '03. New York, NY, USA: ACM, 2003, pp. 111–117. [Online]. Available: <http://doi.acm.org/10.1145/1005140.1005158>