An Aspect-Based Resource Recommendation System for Smart Hotels

Aitor Almeida[1] , Eduardo Castillejo[1], Diego López-de-Ipiña[1], Marcos Sacristán[2], Javier Diego[3]

[1] DeustoTech –Deusto Institute of Technology, Universidad de Deusto
Avenida de las Universidades 24, 48007, Bilbao (Spain)
{aitor.almeida, eduardo.castillejo, dipina}@deusto.es
[2] Treelogic
Parque Tecnológico de Asturias, 30, E33428, Llanera (Spain)
marcos.sacristan@treelogic.com
[3] Logica
Avda. Manoteras, 32, Madrid 28050 (Spain)
javier.diego@logica.com

***Abstract*—The number of resources (services, data, multimedia content, etc) available in Smart Spaces can ver overwhelming. Finding the desired resource can be a tedious and difficult task. In order to solve this problem, Smart Spaces contain much information that can be employed to filter these resources. Using the user context-data available in Smart Spaces can help refining and enhancing the recommendation process, providing more relevant results. To help users finding the most suitable resource we have developed a recommendation system that takes into account both user and resource features and context data like the location or current activity. This recommendation system is flexible enough to be applied to different types of resources and domains. In this paper we describe the resource aspects identified to be used in the recommendation system and how they are combined to create a metric that allows us to select the best resource for each situation.**

Keywords-Smart environments; resource recommendation; context aware; accessibility.

I. INTRODUCTION

Spain is one of the top 5 tourism destinations along France, United States, China and Italy [1]. As a consequence, tourism is an important part of the Spanish economy [2]. In order to maintain this leading position, the hospitality sector must continue evolving and improving, including new technologies and enhancing the user experience. The objective of the THOFU (http://www.thofu.es/) project is to work on the technology that will enable us to create more intelligent and reactive hotels. One of the research areas within the project is the creation of a recommendation system that will enable us to provide the user with the most suitable resources for his current needs and context. In the project's vision resources are anything that a user can consume: apps, hotel services (both physical and virtual), multimedia content, data, etc. The smart hotel must be proactive, helping its users with their needs. In order to do this the smart hotel must know the user preferences, tastes and limitations. It must be capable of analyzing the different aspects that define a resource to offer the most appropriate one to the user. To do this, we have developed an aspect based resource recommendation system. To be able to do this recommendation we have identified the aspects of a resource that can be used to describe it in a smart environment. These aspects take into account both the resource and user features and the current context (as formulated in [3]). Our recommendation system approach has several advantages:

- It is applicable to all the resource types identified in the intelligent hotels domain: digital and physical services, multimedia content and data.
- We evaluate different aspects of the resource taking into account the characteristics of the content, the needs and capabilities of the user and data from the current context. This allows us to create a comprehensive picture of the current situation to recommend the most suitable resource.
- The process can be configured by modifying the weight of each individual aspect in the final metric. This allows us to adapt the recommendation system to specific domains.
- Our system not only analyzes the current situation of the user, it also takes into account what his next actions can be to anticipate future needs.

In this paper, we present the proposed system. In Section 2, we analyze the current state of the art in recommendation systems, in Section 3, we describe the system architecture, in Section 4, we describe some use cases of the proposed system and finally in Section 5, we discuss the conclusions and future work.

II. RELATED WORK

Since the mid-1990s, recommender systems have become an important research area attracting the attention of e-commerce companies. Amazon [4], Netflix (http://www.netflixprize.com/) and Yahoo! Music [5] are widespread examples on making recommendations to its users based on their tastes and previous purchases. Although these systems have evolved becoming more accurate, the main problem is still out there: to estimate the rating of an item which has not been seen by users. This estimation is usually based on the rest of items rated by the current user or on the ratings given by others where the rating pattern is

 ISBN: 978-1-61208-236-3

similar to the user's one. Therefore, the problem consists on extrapolating somehow the utility function (which measures the usefulness of an item to a user) to the whole rating space. This utility function is represented by all the ratings made by the user. This way, recommendation engines have to be able to predict or estimate the ratings of the not yet rated items for users.

The research in this area has, as a result, a different classification based on the way item recommendation is made [6]:

- *Content-based:* recommendations are made just by looking in the history of the already rated items by the user.
- *Collaborative filtering:* past recommendations for users with the same preferences generate recommendations for the current user.
- *Hybrid techniques:* as a combination of content-based and collaborative recommendation approaches.

Content-based systems recommend items which are similar to those that a user rated positively in the past [7]. Shardanand *et al.* [8] state some of the problems of this approach, as the vagueness in the description of an item, which clearly affects the whole system. Items need to have enough descriptive features to enable the recommendation engine to recommend them accurately. The problem is that different items with the same features can be indistinguishable to the system.

Collaborative filtering techniques deal with the concept of similarity between users. The utility of an item is predicted by those items which have been rated by similar users. Sarwar *et al.* [9] defend this approach by defining collaborative filtering as the most successful recommendation technique to date. In [8] a personalized music recommendation system is presented, namely Ringo, which is a social information filtering system which purpose is to advise users about music albums they might be interested in. By building a profile for each user based on their ratings, it identifies similar users so that it can predict if a not yet rated artist/album may be to user's liking. LikeMinds [10] defines a closeness function based on the ratings for similar items from different users to estimate the rating of these items for a specific user. It considers a user which has not already rated the item and a so-called mentor who did it. Introducing two new concepts (horting and predictability) horting is a graph-based technique in which users are represented as nodes and the edges between them indicate their similarity (predictability) [11]. The idea is similar to nearest neighbor, but it differs from it as it explores transitive relationships between users who have rated the item in question and those who have not.

In order to reduce the limitations of previously reviewed methods, hybrid approaches combine both of them [12]. Others have introduced new concepts to this area, such as semantics [13] and context [13].

However, one of the most important improvements in the recommendation systems field is the definition of measures (or aspects) to describe the utility and relevance of the items. Aspects play an important role in data mining, regardless of the kind of patterns being mined [14]. Users' ratings are a good way to trace the interestingness and the relevance of items. Despite of the ratings, there are many measures which allow us to go into these items taking into account the use of them (their consumption) by the users. In other words, we look into the behavior of users for measuring their interestingness for these "items" (for now on we will refer items as resources). From our point of view a resource could be a product, an application or any kind of service (e.g., multimedia, news and weather or connectivity infrastructure services). We have studied several measures from the literature to evaluate those which best fit in our recommender system, such as minimality [15, 16], reliability [17], novelty [18], horting, predictability and closeness [14], and utility [9]. Location is one of the most important measures in many context-aware systems. Several authors has worked in location based recommendation systems [20][21][22]. In these papers, authors use location data captured with GPS and mobile devices to create timely and targeted recommendations for users.

III. System Architecture

To be able to evaluate the suitability of the resources for a given user we have identified a series of aspects that define the identified resource types:

- Physical services: Those services that are used in the real world (e.g the hotel restaurant, pool, gymnasium, etc)
- Virtual services: Services accessed using a device.
- Multimedia content (e.g. video, music, etc)
- Information (e.g. maps, news, etc)

These aspects must be generic enough to be able to use them to describe all the type of resource and expressive enough to capture the different facets of the resources. In the current implementation, we have considered four of them, but we discuss the other ones in the future work section. The four aspects that we currently take into account are the following:

1. Predictability;
2. Accessibility;
3. Relevancy;
4. Offensiveness;

Each one of those aspects is used in the calculation of the suitability value (see Formula 1). The weight of each aspect on the final value can be modified to better adapt the recommendation system to the specific domain of each smart environment. The values of the weights will depend on the requirements of the specific scenarios. For example, if a smart space has a considerable number of users with disabilities, the *accessibility* of the resources will be especially significant. On the other hand, if the scenario is composed by a single space, the *relevancy* will not be as

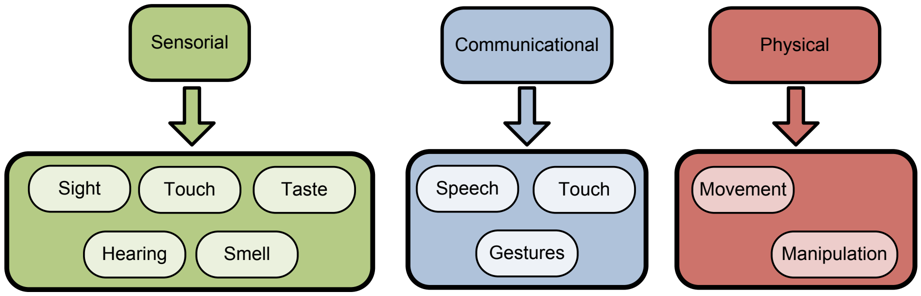

Figure 1. Taxonomy of the user abilities taken into account in the *accessibility* aspect. Disabilities are classified in this three categories.

important as the rest of the aspects. The suitability value is always personalized to a specific user and can change over the time along the preferences of the user.

$$M_{tot} = \sum \omega_i f_i \quad (1)$$

where:

- M_{tot} is the value of the suitability of each resource.
- ω_i is the weight for an aspect.
- f_i is the value of the aspect of a resource. The values of the aspects are normalized

A. Predictability

The first aspect we evaluate is the predictability. This aspect reflects how likely a resource is to be used based on the resources consumed previously. This likeliness is expressed as a probability value between 0 and 1. We use Markov Chains to create the model of the user's resource usage. This model allows us to ascertain patterns in the user behavior. E.g. When one user stays on the hotel his morning routine consists in using the "Press Digest" to recover the headlines of the day, the "Room Service" to order breakfast and the "Transport Service" to call a taxi. With the generated model, we will able to predict that after using the "Room Service" the most probable service to be consumed is the "Transport Service" (see Figure 2).

To build the transition matrix for the Markov Chains, we use the previous history of the user's resource consumption as the training set. This transition matrix can be retrained with the new data recovered from the user with each visit to the hotel, adapting itself to the changes in the user preferences. As we discuss in the future work section, one of the main problems with using Markov Chains is that we only take into account the last consumed resource to predict the next one due to the Markov Property.

B. Accessibility

One of the most important aspects is the accessibility features of the resource. Users of intelligent environments possess a wide variety of abilities (sensorial, cognitive and so on) that must be taken into account to assess the suitability of the resources. Whatever the resource is, users must be able to consume it. We have used the user abilities taxonomy proposed in [19]. We have restricted the user abilities to three groups (see Figure 1):

- *Sensorial abilities:* Those abilities related to the user input.
- *Communicational abilities:* Those abilities related to the user output.
- *Physical:* Those abilities related with the capability of the user to move his extremities.

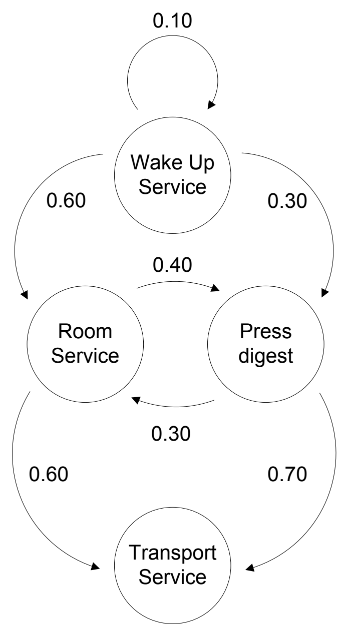

Figure 2. One of the Markov Chains created with the resource consumption data for the *predictability* aspect. Using the created model the recomender system can predict the likeness of one resource to be the next to be consumed.

Each resource has two types of abilities associated, the required and recommended user abilities. If the user does not have one of the required abilities the value of the aspect is automatically set to 0. This is done to reflect the fact that the user can not consume the resource, thus being completely useless for that user. If the user does not have a

 ISBN: 978-1-61208-236-3

recommended ability the accessibility value receives a penalization (see Formula 2).

$$A_{acc} = 1 - \omega|Rec_{not}| \quad (2)$$

where:

- A_{acc} is the accessibility value for the resource.
- ω is the penalization weight.
- $|Rec_{not}|$ is the number of recommended abilities not met by the user.

C. *Relevancy*

This aspect measures the importance of a given resource [20] to the user's current context [3]. For example, a user jogging may be interested in the location of parks and running routes but a user having breakfast in the hotel may be interested instead in the public transports available in the city. One of the main problems we encountered evaluating this aspect was the selection of the context variables. The selected variables must be significant enough to be applicable to all the resource types described previously. We have identified three context variables that meet these requisites. We have analyzed the variables to identify the most common values within the Smart Hotel scenario. In this scenario the most important values are those that are closely related with the hotel, but it also takes into account those were the hotel can offer some service to the users:

1. *User location.* In the tourism domain, we have considered the following locations: client's room, hotel's lobby, hotel's restaurant, hotel's swimming pool, hotel's gymnasium and outside the hotel.
2. *Time of the day.* We have divided the day in twelve periods of two hours.
3. *Current activity.* In the tourism domain we have identified seven activities: sleeping, hygiene routine, eating, exercising, working, shopping and visiting tourist attractions.

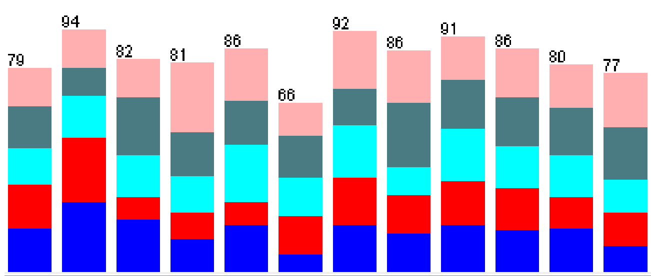

Figure 3. Distribution of the resource consumption in the different periods of time in the used training set for the *relevancy* aspect.

The context information is provided by other modules of the THOFU project that are out of the scope of this paper. Using the usage data recollected from the users we have trained a soft classifier that, given those three context variables, calculates the relevancy of a resource.

For the classifier we have used a nearest neighbor search. KNN (k-nearest neighbor) is a supervised (the training data is labelled), non-parametric (the model does not take a predetermined distribution form but it is in inferred from the data), lazy learning (there is no specific training phase) classification method. KNN assumes that the instances are distributed in a feature space. Since the instances exist in a multidimensional space, there is a computable distance between them. The most commonly used distance is the Euclidean distance. The algorithm takes a user-defined k constant. The instances are classified taken the k nearest training examples in the feature space.

To implement this classifier we have used the libraries included in the Weka framework [29]. We have used LinearNNSearch as the nearest neighbor search algorithm, with a k value of 3 and the Euclidean distance as the distance function.

D. *Offensiveness*

This aspect measures the suitability of a resource based on a rating system. We use the age categories (3, 7, 12, 16 and 18) and the content descriptions (violence, bad language, fear, sex, drugs, gambling, discrimination and online) developed for the PEGI (Pan European Game Information)[31] rating system. To evaluate it we use a similar system that the one used in Section 3.1 to calculate the accessibility, but taking the age categories as required constraints and the content descriptions as the recommended ones.

IV. USE CASE

To better illustrate how the developed system works we will explain how the system works taking two different users as examples. The first user is a 27 year old male with a hearing impairment. The second one is a 6 year old child. The users have five resources available to them in this example: The wake up service (R1), the room service (R2), the press digest (R3), the multimedia system (R4) and the transport service (R5). For this example, the weights for the metric calculation are:

- *predictability* and *relevancy* have a weight of 1
- *accessibility* and *offensiveness* have a weigh of 0.5

We assume that both users are in their rooms and that the wake up service has just been activated by an alarm. The first user uses the Markov Chain model described in

Figure 2. The wake up service and multimedia system both have hearing requirements, but offer alternative means to use them. The first user has not stated any content restriction. The results are shown in Table I.

TABLE I. RESULTS FOR THE FIRST USER

	Predictability	Accessibility	Offensiveness	Relevancy
R1	0.10	0.9	1	0.8
R2	0.60	1	1	0.7
R3	0.30	1	1	0.4
R4	0	0.9	1	0.2

 ISBN: 978-1-61208-236-3

R5	0	1	1	0.3

The second user uses the Markov Chain model described in Figure 4. The user has not any disability, so every resource attains the maximum score in *accessibility*. The press digest has a minimum age category of 7 and it receives a score of 0 in *offensiveness*. The results are shown in Table II.

TABLE II. RESULTS FOR THE SECOND USER

	Predictability	**Accessibility**	**Offensiveness**	**Relevancy**
R1	0.45	1	1	0.2
R2	0.05	1	1	0.1
R3	0	1	0	0.1
R4	0.50	1	1	0.9
R5	0	1	1	0

Using the Formula 1 the recommended resource for the first user will be the room service (R2) in this scenario.

$$M_{tot} = 1 \times 0.6 + 0.5 \times 1 + 0.5 \times 1 + 1 \times 0.7 \qquad (3)$$

In the case of the second user, the selected resource will be the multimedia system (R4).

$$M_{tot} = 1 \times 0.5 + 0.5 \times 1 + 0.5 \times 1 + 1 \times 0.9 \qquad (3)$$

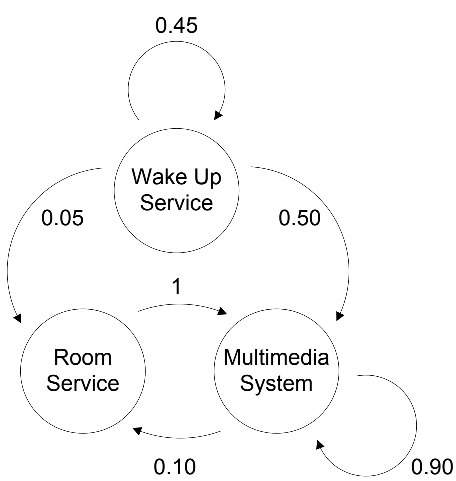

Figure 4. Markov Chain user for the second user

V. CONCLUSION AND FUTURE WORK

In this paper, we have described a resource recommendation mechanism for smart environments based on the evaluation of different aspects of the resources. Our approach provides several advantages:

- The proposed mechanism is generic enough that it can be applied to any type of resource (services, multimedia content, etc). To achieve this we have identified those aspects that are not specific for a given domain or resource.
- We take into account several aspects of a resource, providing a holistic approach to the problem of the resource recommendation.
- The importance of each individual aspect can be tailored for each domain and specific problem modifying their weights in the metric. This allows us to adapt the mechanism to the requirements of specific smart spaces.

One of the problems identified in this approach is the use of Markov Chains to evaluate the predictability aspect. With the use of Markov Chains we only evaluate the current event and not the previous events that preceded it. In order to tackle this problem we plan to explore the use of time series to improve the forecasting algorithm.

We are also analyzing a more extensive set of aspects that will give us a better picture of the evaluated resources. We are currently studying the inclusion of the following aspects:

- *Timeliness* [24]: evaluates how up to date is the information of a resource.
- *Satisfaction* [25][26]: measures the opinion of the users about a resource.
- *Attention* [27][28]: The average number of interactions per time unit with a consumed resource.
- *Closeness* [11]: Evaluates what resources are consumed by similar users.

By adding these new aspects we aim to create more significant resource recommendations that meet better the user's needs. Finally we would like to include in the context data information about the vagueness and uncertainty of the model. To do this we plan to use the ambiguity assessing techniques we described in [30]. This will allow us to model the context more realistically and will improve the overall preciseness of the system.

ACKNOWLEDGMENT

This work has been supported by project grant CEN-20101019 (THOFU), funded by the Spanish *Centro para el Desarrollo Tecnológico Industrial (CDTI)* and supported by the Spanish *Ministerio de Ciencia e Innovación*.

REFERENCES

[1] Interim Update. UNWTO World Tourism Barometer (UNWTO). URL: http://www.unwto.org /facts/eng/pdf/barometer/UNWTO_Barom11_iu_april_excerpt.pdf, retrieved: July, 2012.

[2] Economic report. Bank of Spain. URL: http://www.bde.es/informes/be/boleco/coye.pdf, retrieved: July, 2012.

[3] Dey, A.K. Understanding and using context. Personal and ubiquitous computing, vol 5, no. 1, pp 4-7, 2001.

[4] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," Internet Computing, IEEE, vol. 7, no. 1, pp. 76–80, 2003.

[5] P. L. Chen *et al*., "A Linear Ensemble of Individual and Blended Models for Music Rating Prediction.". KDDCup 2011 Workshop, 2011.

[6] G. Adomavicius and A. Tuzhilin, "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, vol. 17, no. 6, pp. 734–749, 2005.

[7] D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles, "Collaborative filtering by personality diagnosis: A hybrid memory- and model-based approach," in Proceedings of the 16th conference on uncertainty in artificial intelligence, 2000, pp. 473–480.

[8] U. Shardanand and P. Maes, "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," 1995.

[9] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," in Proceedings of the 10th international conference on World Wide Web, 2001, pp. 285–295.

[10] D. Greening, "Building consumer trust with accurate product recommendations," LikeMinds White Paper LMWSWP-210-6966, 1997.

[11] C. C. Aggarwal, J. L. Wolf, K. L. Wu, and P. S. Yu, "Horting hatches an egg: A new graph-theoretic approach to collaborative filtering," in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 201–212.

[12] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining content-based and collaborative filters in an online newspaper," in Proceedings of ACM SIGIR Workshop on Recommender Systems, 1999, p. 60.

[13] S. Kim and J. Kwon, "Effective context-aware recommendation on the semantic web," International Journal of Computer Science and Network Security, vol. 7, no. 8, pp. 154–159, 2007.

[14] L. Geng and H. J. Hamilton, "Interestingness measures for data mining: A survey," ACM Computing Surveys (CSUR), vol. 38, no. 3, p. 9, 2006.

[15] B. Padmanabhan and A. Tuzhilin, "Small is beautiful: discovering the minimal set of unexpected patterns," in Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, 2000, pp. 54–63.

[16] Y. Bastide, N. Pasquier, R. Taouil, G. Stumme, and L. Lakhal, "Mining minimal non-redundant association rules using frequent closed itemsets," Computational Logic—CL 2000, pp. 972–986, 2000.

[17] P. N. Tan, V. Kumar, and J. Srivastava, "Selecting the right interestingness measure for association patterns," in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 32–41.

[18] S. Sahar, "Interestingness via what is not interesting," in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 332–336.

[19] A. Almeida, P. Orduña, E. Castillejo, D. López-de-Ipiña, M. Sacristán. Imhotep: an approach to user and device conscious mobile applications. Personal and Ubiquitous Computing. vol 15, no. 4, pp 419–429 January 2011.

[20] V.W. Zheng. Y .Zheng, X. Xie, Q. Yang. Collaborative location and activity recommendations with gps history data. Proceedings of the 19th international conference on World Wide Web. 2010.

[21] Y. Zheng, L. Zhang, Z. Ma, X. Xie and W.Y. Ma. Recommending friends and locations based on individual location history. ACM Transactions on the Web, vol 5, no. 1, pp 5. 2011.

[22] V.W. Zheng, B. Cao, Y. Zheng, X. Xie and Q. Yang. Collaborative filtering meets mobile recommendation: A user-centered approach. Proceedings of the 24th AAAI Conference on Artificial Intelligence. 2010

[23] Leo Pipino, Yang Lee, and Richard Wang. Data Quality Assessment. Communications of the ACM, 45:211-218, 2002. PLW02

[24] Ballou, D.P., Wang, R.Y., Pazer, H. and Tayi, G.K. Modeling information manufacturing systems to determine information product quality. Management Science 44, 4 (1998),462–484.

[25] Bo Pang, Lillian Lee. Opinion Mining and Sentiment Analysis. Journal of Foundations and Trends in Information Retrieval vol 2, no. 1-2, 2008.

[26] Beatson, A. and Coote, L.V. and Rudd, J.M. Determining consumer satisfaction and commitment through self-service technology and personal service usage. Journal of Marketing Management, vol 22, no. 7-8, pp 853-882, 2006.

[27] S. Kim and J. Kwon, "Attention-Based Information Composition for Multicontext-Aware Recommendation in Ubiquitous Computing," Smart Sensing and Context, pp. 230–233, 2006.

[28] T. H. Davenport, "May we have your attention, please?," Ubiquity, vol. 2001, no. June, p. 3, 2001.

[29] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten. The WEKA Data Mining Software: An Update. SIGKDD Explorations, Volume 11, Issue 1. 2009

[30] A. Almeida and D. López-de-Ipiña. Assessing Ambiguity of Context Data in Intelligent Environments: Towards a More Reliable Context Managing System. Sensors. vol 12, no. 4, pp 4934-4951. April 2012.

[31] Pan European Game Information (PEGI). Available online: http://www.pegi.info/, retrieved: July, 2012.