

Why Numbers, Invites and Visits are not Enough: Evaluating the User Experience in Social Eco-Systems

Lynne Hall, Colette Hume
Department of Computing, Engineering and Technology
University of Sunderland
Sunderland, SR6 0DD, United Kingdom
{lynne.hall, colette.hume} @sunderland.ac.uk

Abstract — Social eco-systems are often evaluated through quantitative data that is automatically logged and analysed. However, where the user’s experience of social eco-systems is evaluated, more explicit intervention approaches are typical, with questionnaires, focus groups and user testing widely used, directly asking the user about their experience. User experience evaluation thus ruptures the social eco-system, occurring as a separate, discrete activity outside of that system. In this paper, we propose that evaluation should be part of the social eco-system adding value to the user experience. We outline an evaluation approach that has been applied within games-based learning environments where the evaluation is seamlessly embedded. We briefly outline our approach to generating and analyzing data highlighting its potential for social eco-system evaluation.

Keywords-evaluation; user experience; analysis of user-generated content.

I. INTRODUCTION

It is widely recognized that the impact of social eco-systems requires further consideration, with relatively few studies or empirical investigations. Social eco-systems typically involve enjoyable and often affective interactions within a user-chosen context. The user’s interaction focus is primarily recreation, enjoyment or problem solving in relation to a social need. Yet, how can we evaluate or understand the impact of that interaction on an individual or societal level? And more, if we do try to evaluate it, can we do this without having an experimenter effect, even if that “experimenter” is an anonymous on-line survey.

Users are only vaguely aware and in general don’t seem to care about the collection of usage statistics. Thus, statistics can be endlessly calculated relating to the number and frequency of visits, invites, postings and so on, without any impact on the user. However, evaluating the user experience is more challenging, requiring conscious user input, rather than logging of actions.

Unlike the integrated usage data collection, the user experience evaluation of social eco-systems is typically a separate, discrete activity to the main use of the system, with questionnaires, focus groups and user testing widely used. User experience evaluation thus changes the dynamic of the

social eco-system, placing the user in the role of evaluator rather than social network member.

In this paper, we propose an alternative to this discrete, separate approach to user experience evaluation. Instead of separating out evaluation and changing the role of the user, we have developed an approach that enables us to evaluate the user experience without users being aware that they are taking part in an evaluation. This approach has considerable relevance to the evaluation of social eco-systems, meeting two key success factors for social networks:

- Evaluation should be invisible and should have no (as achieved with usage statistics) or a positive impact on user activities
- Add-ons (e.g. evaluation instruments) to the social eco-system must be integrated and add value to the user experience

In this paper, we briefly outline our approach to the generation of evaluation content and discuss our proposed approach to the analysis of this content. Our key focus is how to mask the evaluation experience so that the user is unaware of their evaluation input whilst generating data useful to an interdisciplinary research and design team. This approach has been successfully applied and we believe that it offers potential for other developers and researchers to evaluate social eco-systems. Section 2 briefly discusses social eco-system evaluation, highlighting the focus on commercial factors and the relevance of these to user experience evaluation. Section 3 discusses our approach to user experience evaluation, outlining our approach and its application to two systems. Section 4 discusses our approach and considers its potential for evaluating social eco-systems. Section 5 concludes that this approach has considerable relevance to supporting and improving the user experience of evaluation.

II. EVALUATING SOCIAL ECO-SYSTEMS

There has been a massive growth in commercially supported social eco-systems. The marketers, quite rightly, recognize that supporting an on-line community will increase brand loyalty and sales. Through allocating significant resources to on-line activity, some companies have established high quality, effective social eco-systems, with significant user presence. The purpose of these social eco-systems is to enable companies to achieve their business

goals. Thus, in the evaluation of such commercially derived social eco-systems the evaluation issue is not really user experience and social impact, rather it is the company's Return On Investment (ROI). This ROI includes the social eco-systems impact on: developing brand loyalty, thought leadership, reducing operating costs, optimizing marketing budgets, and increasing profits [1].

With the aim of demonstrating ROI, much of the evaluation in social eco-systems is achieved using logged user interactions. For example, the number of invites made by a user; frequency of postings; and number and type of interactions within the social eco-system. There are many tools available to log and analyse user interactions, with such functionalities increasingly provided as standard in site development products. However, whilst tools can be used as a basis to calculate a range of quantitative measures such as visits, social graph, social surface area, etc. their insight into the direct user experience is limited. Whilst such numerical data can enable us to determine the strength, sustainability and growth potential of the social eco-system, it does not allow us to explore the user experience itself.

There are considerable challenges for user experience evaluation of social eco-systems, with users often geographically dispersed and having limited real world interactions. In response to this, techniques have been developed for both virtual and real world evaluations. However, the majority of these require additional user input, often with the user role changing from member, player, commentator, etc. to a critic, tester or evaluator.

Whilst engagement in user experience evaluation can offer positive benefits to participants, for example, early access to new features, input to development, status within the network, etc., many users choose not to participate in evaluations. Thus, unless participation in evaluation activities is mandated (e.g. in a fiat system [2]), the participants self-select thus providing only a partial view of the user experience of the social eco-system. Further, where participation in evaluation activities is mandated, users can view evaluation as a burden [3].

In considering the evaluation of the user experience in a social eco-system, it is not the issue of usability that is key. There are a whole variety of half-hearted attempts by companies and organisations to create social eco-systems. From these, we know that if the usability is poor that unless the environment is incredibly compelling, then users will go elsewhere. Instead, it is the user's personal, social and emotional experience that requires evaluation to enable us to explore the impact of social eco-systems.

III. EMBEDDING EVALUATION IN THE USER EXPERIENCE OF SOCIAL ECO-SYSTEMS

Our approach to evaluation has been developed within the EU FP6 eCIRCUS [4] and FP7 eCUTE [5]. Both projects have focused on technology enhanced learning for significant social issues, including bullying and intercultural conflict. In this paper, we discuss our evaluations with the ORIENT [5] and MIXER [6] showcases, outlining our

approach and highlighting the potential for its use with other social eco-systems.

Our research has focused on evaluating a specific type of social eco-system: technology enhanced learning through interaction in intelligent computer assisted role-play environments. In our experiences of designing, developing and evaluating our showcase applications, we have dramatically changed our approach to evaluation. Rather than evaluation being conducted as a discrete, separate activity to the interaction, we now add value through seamlessly embedding evaluation into the user experience. The impact of this is that users are unaware they are taking part in an evaluation. In addition, the results from this evaluation have been of considerable use to the interdisciplinary development team.

To enable us to evaluate our showcases, users are actively engaged in the individual and communal generation of real world artefacts and digital assets. Critical to the success of our approach is for users to be aware of, and participate in the social eco-system provided through our environment. We artificially create a temporary social eco-system for a specific showcase and its participating users. Whilst we have to stimulate users into creating assets, in many social eco-systems a plethora of such user-generated content exists or could easily and enjoyably be developed meeting the requirements of the evaluation and improving the user experience.

However, having extensive data or content is insufficient without a viable analysis approach. Analysing the content is complicated by a multiplicity of formats and the challenges offered by non-textual assets. Our evaluation approach uses a range of techniques and tools for content analysis, with approaches derived from information retrieval research transforming the content into usable data.

The following examples briefly outline our approach to generating and analyzing user experience data.

A. ORIENT: Seamlessly embedding evaluation into the user experience

ORIENT provides users with an intelligent computer assisted, semi-immersive, graphical role play environment depicting an imaginary culture, the 'Sprytes.' It is aimed at teenagers and young adults who interact in groups of 3, taking roles in Space Command (a benevolent United Nations type of organization with a galactic focus) with the goal of helping the Sprytes to save their planet from imminent destruction. ORIENT's learning focus is cultural understanding and sensitivity.

The characters, the Sprytes, inhabiting this world are autonomous agents, based on an extension of the FATiMA agent architecture [7]. Emotional appraisal is based on the OCC cognitive theory of emotions [8] extended by incorporating aspects of a needs driven architecture, PSI [9]. To enable cultural adaptation of the agents, Hofstede's cultural dimension values were added to the agent minds for the culture of the character; cultural specific symbols; culturally specific goals and needs, and the rituals of the culture [10].

Users interact with the Sprytes using a Wiimote to provide gestures and speech recognition of character names. They interact with the ORIENT world using a scanner phone with an RFID reader. Additionally, the users are provided with the ORACLE (Onboard Resource Agent - Cultural and Liaison Engagement), a mobile phone based embodied conversational agent whose role is to support the users in their interaction. Figure 1 provides an overview of ORIENT’s main components. At the core of the system is the virtual world model that is presented to the user as 3D graphics on a large screen, in front of which the users interact with ORIENT as a group.

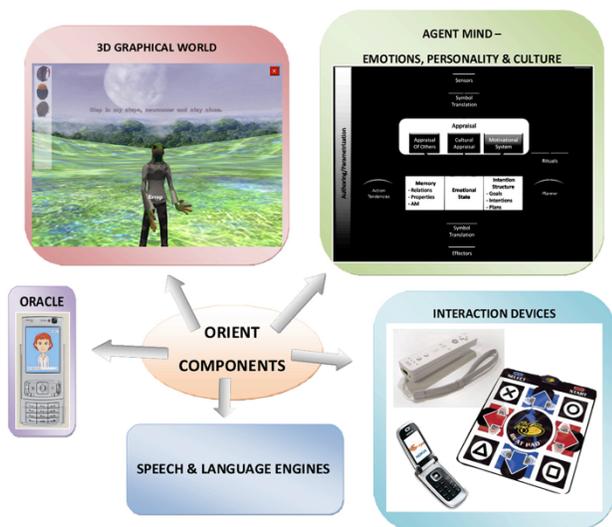


Figure 1. ORIENT Overview

Developed as part of an interdisciplinary project, the evaluation aimed to investigate the effectiveness of ORIENT in fostering cross-cultural acceptance through the promotion of collaborative practices and the appreciation of similarities and differences between cultures. From the technical perspective, evaluation focused on the coherence and comprehensibility of the narrative; the believability and credibility of the agents that underpin the characters; and participant engagement with the cultures of ORIENT and the Sprytes themselves. With the interaction approach, we focused on evaluating the participant’s views of the impact of unusual interaction devices and mechanisms, focusing on device usability and user satisfaction with unusual interaction mechanisms. This resulted in a wide range of purposes and instruments required for the evaluation.

Even though we needed users to participate in an extensive evaluation, our goal was for players to have only one consistent experience that of being a player in a role play game. To achieve this we transformed traditional and/or well established data gathering instruments into ‘in role’ counterparts. These were then embedded into the role play and reinforced with supporting artifacts. Each instrument was given archetypal branding (adding value to the role play context) and an age appropriate format and aesthetic (meeting user expectations), see figure 2. The resulting

battery of piloted instruments aimed to add maximum value to the over-arching role playing game while collecting key evaluation data to help developers assess the user experience from a number of theoretical perspectives.

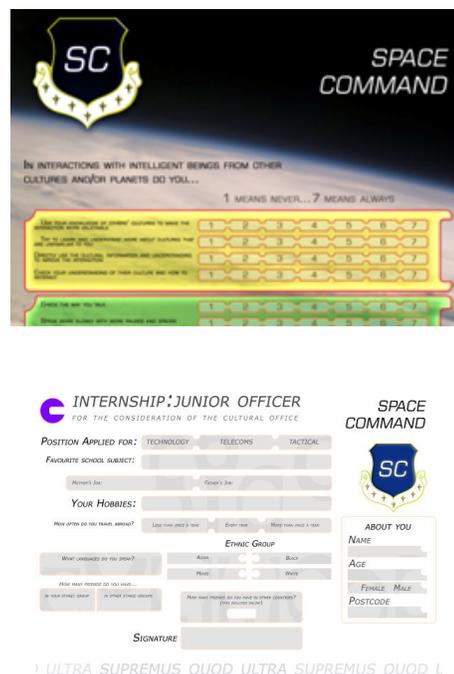


Figure 2. Evaluation Instruments

This approach was very successful in generating data from users about their experience, with the interesting side-effect that users were completely unaware where the game stopped and the evaluation started. The evaluation instruments and activities are effectively seamless and thus data captured in a way that is invisible for the user. Rather than the evaluation instruments and supporting artifacts adding a burden to the user, they seemed instead to enhance the game, actually increasing the immersion and enjoyment of the users. The data and content produced through the user interactions was analysed using qualitative and quantitative analysis techniques and are further discussed in [11].

B. MIXER: Opinion and sentiment: approaches to analyzing user generated content

With ORIENT, the majority of the user-generated content was achieved through specially prepared instruments many of which were hard copy. With the ongoing development of our evaluation approach, we are focusing on the generation of digital assets. Our exploration of the generation and analysis of digital assets is currently focused on MIXER [6]. This application aims to provide 9-11 year olds with classroom-based, technology enhanced learning experiences related to cultural conflict. This context for MIXER is provided by Hide & Seek where participants may be characters or other users and where conflict is typically a result of rule misunderstandings, based on Hofstede's cultural dimensions [10]. Figure 3 provides some frames outlining the MIXER narrative.

Our evaluation is focused both at children and teachers as achieved through their interactions with MIXER and their discussing of these experiences. The evaluation is seamlessly embedded into the experience of the application, right from the initial design stage. For example, the frames in figure 3 have been generated as a comic book. Into this comic book (which represents the application for the users) we have embedded traditional questionnaires that have been morphed into quizzes and mini-games.

In addition, the comic book is supplemented by an on-line experience, where the users will engage in the generation of blogs, digital AV & photo albums and participation in a tailored social network. Two complementary social networks are used, one for the teachers and the other for the child users of MIXER.

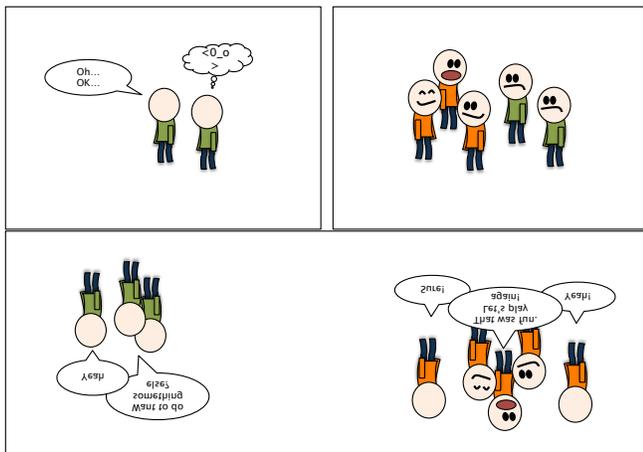


Figure 3: Frames from the MIXER storyboard

During the user interaction, a main focus is gaining an appreciation of the user's Theory of Mind with users prompted with a range of questions which result in the user producing freeform text in the interaction. In addition to this direct input during the interaction users are also involved in generating content in relation to their experience.

An initial study, involving three groups of 9-11 year old children (10 in each group) has recently been conducted to establish user responses to embedding evaluation in MIXER.

The three conditions were provided in separate locations. This separation was to ensure that children of one group were not aware of the other conditions' activities. The three evaluation conditions were:

- **Direct evaluation:** children were provided with a non-interactive comic book (just for reading); a work book composed of a series of activities related to the comic book and specifically activities related to Theory of Mind; and a set of questionnaires related to the comic book, attitudes to culture, in-group / out-group, understanding of cultural dimensions
- **Hybrid evaluation:** children were provided with an interactive comic book and asked to write / draw responses (thus incorporating Theory of Mind activities). The workbook (without the Theory of Mind activities) and questionnaires were given as a single item, with the questionnaires embedded in the workbook.
- **Seamless evaluation:** children were provided with a single artefact incorporating the interactive comic book, the workbook activities and the questionnaires. The questionnaires were modified and presented as quizzes and activities using age appropriate aesthetics. The activities and questionnaires were placed throughout the comic book, replicating the approach of magazines for 9-11 year olds.

Observations during the use of MIXER highlighted that children in the Seamless Evaluation condition engaged for longer and were highly engaged in the workbook and evaluation activities. Children in this condition required very little input or encouragement from the adults present and worked steadily through the entire artefact. More questions and issues were raised in the other conditions, particularly in relation to completion of the questionnaires. In the direct evaluation condition, children were not particularly interested in completing the questionnaires and spent significantly less time with the comic book and Theory of Mind activities, then those in the Hybrid or Seamless Evaluation conditions.

We are currently engaged in analyzing the data generated during this initial MIXER study. Early results indicate that the results from using different versions of the questionnaires are relatively similar across the conditions. This is an expected outcome illustrating that although instruments are modified they are essentially collecting the same data. However, in line with their greater engagement, children in the Seamless Evaluation condition wrote and drew more within the comic book (and the embedded Theory of Mind activities) than the other conditions. Our initial results appear to indicate that improving the user experience of evaluation results in greater user engagement. Current work focuses on further analyzing our data, particularly in relation to the impact of embedding the Theory of Mind within the Comic book.

With MIXER, we are now focusing on the analysis of freeform text and digital (e.g., audio, video, photos)

contributions. Our key aim in evaluating these user-generated content is to determine personal, social and emotional user experience. As such, we are particularly looking for opinions and affective views within user-generated content.

There is current considerable interest in the evaluation of “opinionated content” such as discussion groups, blogs, tweets, video postings and other methods where people express their views online. Through evaluating relevant user-generated content, it is obvious that companies can gain consumer feedback about their own and competitor’s products, thus avoiding the need to conduct surveys, organise focus groups or employ external consultants [12].

A considerable number of statistical measures can help analyse text and automation tools. Through the use of semi-automated methods, we will be analyzing user generated content provided in the MIXER social eco-system. We are currently investigating a range of methods, aiming to find the most appropriate analysis approach for our evaluation purposes. These interdisciplinary evaluation purposes are quite broad, relating to educational, psychological, socio-cultural, interaction and technical goals. Methods we are investigating include:

- Use of base and comparative polar words (e.g., base: “bad”, comparative: “worse”) enabling the use of statistical measures (e.g., [13]).
- Seed words and connectives such as AND, OR, BUT, or HOWEVER are being used to find related or contrasting words, as in [14].
- Clustering techniques, such as Factor Analysis are being used to identify word and opinion clusters.
- Named entity recognition (as applied in ontology generation) will be applied, aiming to support co-reference resolution, for example – a pronoun such as “it” might refer to “the game”, “MIXER”, “the computer,” etc.
- Synonym grouping will be facilitated using *SenitWordNet* (as used by [15])

A key benefit of using sentiment analysis is that it can be used to convert natural language texts into structured data, that can then be stored and manipulated in a database. We will use Liu’s approach [12] and store user generated content as a quintuple:

- Object (product, person, event, organisation, topic),
- Feature of that object
- Polarity of the opinion of the holder on that feature of that object
- Opinion holder
- Time when the opinion made by opinion holder

This data can be both analysed statistically and represented visually, supporting a greater understanding of the data. Although we have just begun applying this approach to our analysis of MIXER, early investigations suggest that this will provide a powerful addition to our evaluation approach.

IV. DISCUSSION

The Six Benchmarks for Digital Marketing Strategy [16] have been developed to evaluate the potential effectiveness of social media on ROI:

1. Goal - What is the targeted goal of your advertisement, social media program or campaign?
2. Engage – How effective is the message in attracting or involving your target market?
3. Relationship – Did the message stimulate the target to feel trust or common interests?
4. Value –Does the product or service and related message communicate added benefit for the individual, organization or company?
5. Action- Does the message move you to act?
6. Synergize- Is the tool an add-on to current marketing efforts or is it integrated into the campaign?

Although such benchmarks identify plentiful questions and issues, there is little information about how systems can be evaluated against them. Whilst usage stats will answer some issues, clearly, user experience data has to be both generated and analysed to permit evaluation against these benchmarks.

In this paper, we have proposed an approach to the generation and analysis of user generated content. Our approach differs from many current user experience evaluation approaches. Through focusing both on reducing the visibility of evaluation participation and on adding value through evaluation our approach gains useful data whilst either having no or a positive affect on the user.

Our approach to gathering user experience data involves the use of existing user input formats (e.g. blogs, postings, tweets) and the creation of add-ons (e.g., questionnaires represented as quizzes, mini-games, etc.). Our users are consistently unaware that they are taking part in an evaluation. Results have highlighted that users view the evaluation experience positively, seeing it as a value add rather than a negative. In addition, the interdisciplinary project team have gained results and evaluation data that have been relevant and useful.

Within our approach, we are gathering data in two ways. Firstly, through crafting customized quizzes and embedding questions (from existing traditional questionnaires) in interactions and entertaining activities. And secondly, through viewing user generated content as a primary source of evaluation material. Where possible we avoid technology learning and thus use popular formats, Facebook has already trained most of our users.

Sentiment analysis and opinion mining offer considerable potential for the analysis of user generated content in the evaluation of any social eco-system. Semi-automated approaches can greatly increase the speed of data refinement and analysis. The use of such approaches also provides the data in a format that is relatively easy to visualize, thus allowing greater understanding by development teams and stakeholders.

Related work focuses on the evaluation of AV and photographic content. With photography we are exploring indexicality to support evaluation [17]. With both photographs and AV content, the critical issue is how to transform the content into analyzable outputs. Initial results suggest that the labels and descriptions frequently generated by users along with non-textual postings may contain sufficient content to analyse the AV without requiring additional data refinement. To further investigate we are exploring the use of meta-tagging, to enable us to compare results from further content refinement with the use of user generated labels and descriptions.

V. CONCLUSIONS

It is possible to create a user experience evaluation that can be completely embedded within a social eco-system. Evaluation instruments and approaches can be crafted to enhance rather than detract from the social eco-system experience. Sentiment analysis and opinion mining transform user generated content into a highly valuable and analyzable data source. The use of this approach allows user experience evaluation data to be gained and analysed as invisibly as usage statistics.

VI. ACKNOWLEDGMENTS

This work was partially supported by European Community (EC) and is currently funded by the ECUTE project (ICT-5-4.2-257666). The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

VII. REFERENCES

- [1] Kaufman, I., (2010) 5 Steps to Evaluation Social Media ROI, Social Media Today, Available at: <http://socialmediatoday.com/irakaufman1/111688/5-steps-evaluating-social-media-roi> Accessed 25-07-2011
- [2] Hinchcliffe, D., (2010) The K-factor Lesson: How Social Ecosystems Grow (Or Not), Enterprise Irregulars, Available at: <http://www.enterpriseirregulars.com/9499/the-k-factor-lesson-how-social-ecosystems-grow-or-not/> Accessed 25-07-2011
- [3] Sapouna, M., Wolke, D., Vannini, N., Watson, S., Woods, S., Schneider, W., Enz, S., Hall, L., Paiva, A. and Aylett, R. (2010) Virtual Learning Intervention to Reduce Bullying Victimization in Primary School: A Controlled Trial, *Journal of Child Psychology and Psychiatry*, 51(1), pp. 104-112.
- [4] <http://www.e-circus.org> Accessed 25-07-2011
- [5] Hall, L., Jones, S., Aylett, R., André, E., Paiva, A., Hofstede, G.-J., Kappas, A., Nakano, Y., and Nishida, T. (2011) Fostering Empathic Behaviour In Children And Young People: Interaction With Intelligent Characters Embodying Culturally Specific Behaviour In Virtual World Simulations, in *INTED 2011 (International Technology, Education and Development Conference)*, Valencia, Spain, 7-9 March, 2011.
- [6] Hall, L., Hall, M., Hodgson, J., Nazir, A. and Lutfi, S. Games based learning for Exploring Cultural Conflict, in *AISB 2011 Symposium: AI & Games*, York, 6-7 April, 2011.
- [7] Dias, J., and Paiva A. (2005) Feeling and Reasoning: a Computational Model. *12th Portuguese Conference on Artificial Intelligence, EPIA (2005)*, pp. 127-140.
- [8] Ortony, A., Clore, G. L., and Collins, A. (1988) *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge, UK.
- [9] Dörner, D. (2003). *The mathematics of emotion. The Logic of Cognitive Systems: Proceedings of the Fifth International Conference on Cognitive Modeling*, Springer, pp. 127-140.
- [10] Hofstede, G., Hofstede, G.J. and Minkov, M. (2010) *Cultures and Organizations, Software of the Mind, Intercultural Cooperation and It's Importance for Survival (3rd Edition)* McGraw-Hill Books, New York
- [11] Aylett, R., Paiva, A., Vannini, N., Enz, S., Andre, E. and Hall, L. (2009) But that was in another country: agents and intercultural empathy, *8th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2009)*, Budapest, Hungary, May 10-15, pp. 329-336
- [12] Liu, B. (In press). Sentiment Analysis and Subjectivity. To appear in *Handbook of Natural Language Processing, Second Edition*, (editors N Indurkha and F Damerau).
- [13] Turney, P. (2002). Thumbs up or thumbs down. Semantic orientation applied to unsupervised classification of reviews. *Proc. ACL 2002*, pp. 417-424.
- [14] Hatzivassiloglou, V. and McKeown, K. (1997). Predicting the semantic orientation of adjectives. *Proc. Joint ACL/EACL conference*, pp. 174-181.
- [15] Devitt, A. and Ahmad, K. (2007). Sentiment Analysis in financial news: a cohesion-based approach. *Proceedings of ACL*, pp. 984-991.
- [16] Kaufman, I. (2011) Six Benchmarks for Digital Marketing Strategy, Social Media Today, Available at: <http://socialmediatoday.com/patsystemart/262270/gervas-6-benchmarks-digital-marketing-strategy> Accessed 25-07-2011
- [17] Jones, S. and Hall, L. (2011) The photograph as a Cultural Arbitrator in the Design of Virtual Learning Environments for Personal and Social Education for Children and Young People. *Electronic Visualisation and the Arts 2011*, London, 6-8 July, 2011.