Journal of Universal Computer Science, vol. 14, no. 19 (2008), 3114-3126
submitted: 23/7/08, accepted: 29/10/08, appeared: 1/11/08 © J.UCS

Displaying Pictures according to the Songs Being Played

Jesús Ibáñez
(Pompeu Fabra University, Barcelona, Spain
jesus.ibanez@upf.edu)

David García
(Pompeu Fabra University, Barcelona, Spain
david.garcian@upf.edu)

Oscar Serrano
(Pompeu Fabra University, Barcelona, Spain
oscar.serrano@upf.edu)

Abstract: This paper presents Musimage, a novel system which displays pictures according to the songs being played at the same time. By using the interface the user selects the songs to be played, but the pictures are automatically selected. For each song to be played, the system selects a set of pictures, according to various criteria corresponding to some features of the song. In this sense, the pictures to be shown are, metaphorically, triggered by the songs.

Key Words: intelligent user interface, indirection of information, ambient intelligence
Category: H.1.2, H.5.1, H.5.2, H.5.5

1 Introduction

Music triggers recollections. By listening to a particular song, we can remember events that happened and feelings we had when we used to listen to that song in the past. Some songs do not evoke concrete instants but epochs in our life. Songs we used to listen to in 1994 will probably trigger some recollections from that year or the next few years. This is the idea that led us to design and develop Musimage, the system presented in this paper. Our original idea was to design a user interface which on the one hand accompanies the user in this recollection process, and on the other hand is able to illustrate the song.

During the last few years, we are starting to know better some aspect of the power of evocation of music. Thus, in [LIAFLine 2004] it is said: ”For people with Alzheimer’s disease, as for all of us, music is also a way to trigger forgotten memories and associations. Researchers have found that music is processed in some of the same areas of the brain that also store memories, and a familiar song can take one back to the feelings experienced when it was first heard”.

Moreover, a study about long-term memory for popular music is described in [Schulkind et al. 1999]. The subjects of the study listened to excerpts of popular songs (drawn from across the 20th century) and they had to give emotionality

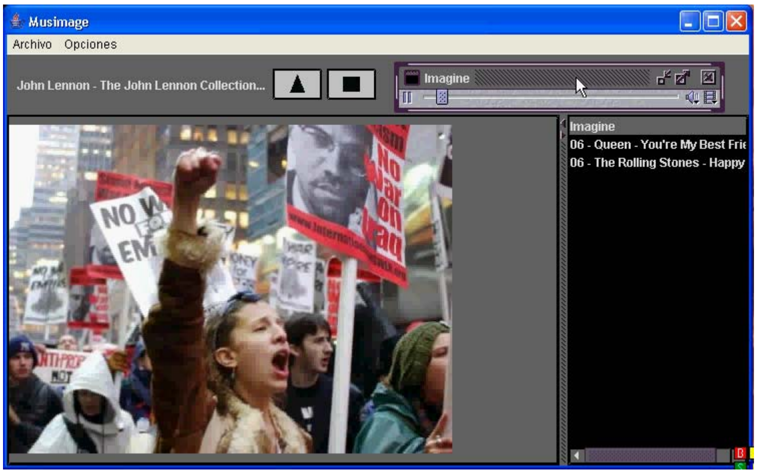
Musimage
Archivo
Opciones
John Lennon - The John Lennon Collection...
Imagine
Imagine
06 - Queen - You're My Best Frie
06 - The Rolling Stones - Happy

Figure 1: User interface

and preference ratings to the songs. The results showed that older adults' emotionality ratings were higher for songs from their youth. They also remembered more about these songs. However, the stimuli failed to cue many autobiographical memories of specific events.

On the other hand, the value of pictures as stimulus for recollecting is undeniable. As stated in [de Ruyter 2003], "photographs are irreplaceable representatives of memories that people have. They are the undisputed number one in the ranks of important objects".

Therefore, it makes sense to think of a system which augments the power of evocation of music by illustrating the user experience with pictures. Musimage, the system presented in this paper, displays pictures according to the songs being played at the same time. By using the interface (see figure 1), the user selects the songs to be played, but the pictures are chosen automatically. For each song to be played, the system selects a set of pictures, according to various criteria corresponding to certain features of the song (namely lyrics and year). The topics of a song are automatically obtained from its lyrics through a categorization process. These semantic features (topics and year) are then used as cues for collecting pictures from both the user's personal picture collection and picture servers on the Internet.

The rest of the article is structured as follows. We first describe the architecture and overall functioning of the system. Following this, we detail the mechanisms we have designed to select pictures depending on the songs being played. Then, we describe several applications and scenarios where the system can be used. Finally, we include the conclusions and future work. References to related work are included through the whole paper.

2 Architecture and overall working

The work presented in this paper is included in a more generic line of research where we are exploring, in a more general way, the indirection of information. We study how the user's actions can be translated, indirectly, into other actions beyond the original user's intention. This connects with the ubiquitous computing [Weiser 1991][Weiser 1993] and ambient intelligence [Marzano and Aarts 2003] paradigms. It also connects with both Mitchel's [Mitchell 2000] and Dertouzos' [Dertouzos 2001] visions. All of these paradigms and visions share a scenario where everyday objets contain a digital component which connect them to networks which, all together, constitute a network of invisible information which surrounds us and mixes itself with the physical environment we knew so far. With this horizon, we are exploring mechanisms which permit that the user's interaction with this network can have adequate consequences beyond the essentially expected by the user.

Thus, although in the system we present in this paper we deal with songs (as triggers) and pictures (as triggered medias), when designing the architecture we tried it to be flexible and extensible, so that it is able to work with other media and interrelations among them. As a result of these considerations and requirements, the system architecture is a multi-agent system [Ferber 1999]. More concretely, the system has been developed with the JADE (Java Agent DEvelopment) framework [Bellifemine et al. 2007]. Designing the architecture as a multi-agent system facilitates the separation of different functionalities into different components (agents). It also facilitates the addition of new agents providing new functionalities. Thus, in the application presented in this paper both the song player and the picture visualizer are two different agents and, although accessible from the same user interfaces, they are easily separable. Therefore, the system could, for instance, display pictures in a digital picture frame by taking into account the music we are listening to in our laptop.

2.1 Architecture

Figure 2 shows the overall architecture of the current system. It is based on a network of brokers or mediators. An agent obtains the services it requires by requesting them to its local broker. When a local broker is not able to solve the requested task, it requests the task in turn to the network of external brokers in a transparent way for the client agent. In this sense, both the song player and the picture visualizer are agents connected to a local broker. The song player informs the broker about the songs it plays and is going to play. The picture visualizer request the broker pictures to visualize according to the actions of the

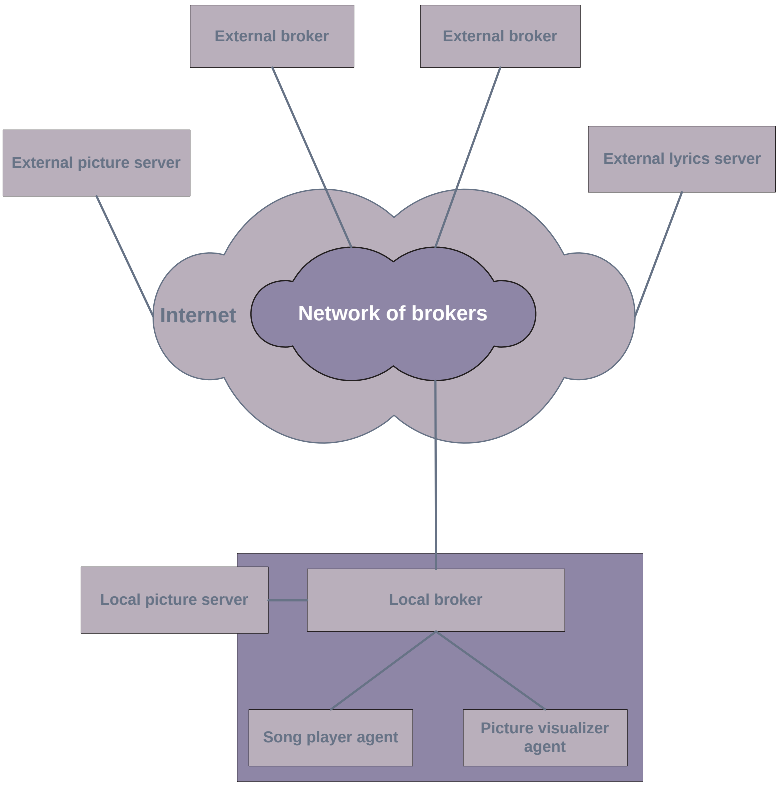

Figure 2: Overall architecture

song player agent. The broker is the agent which, through a series of processes which we describe later, decides which pictures are triggered by the songs played by the song player, and informs the picture visualizer about that. In order to achieve its objectives, the broker requires, in turn, accessing several services: access to servers of lyrics on the Internet, automatic categorization of lyrics, and access to servers of pictures. If the local broker is unable to carry out any of these tasks, it will ask the external brokers for its completion.

3 Overall working

The system employs text categorization techniques to assign a set of weighted topics to song lyrics. These weighted topics are then contrasted with the picture topics. Therefore, the system requires two previous steps for it to work properly:

- On the one hand, it needs the categorizer of song lyrics to be previously trained with a representative set of lyrics.

- On the other hand, it requires that the user's personal collection of pictures is annotated with keywords (topics) and a date.

The training of the categorizer is not a big problem, as it can be done just once. Moreover, it is possible to reuse a categorizer trained by other users. In our particular case, we trained a categorizer able to classify lyrics into three categories: love, friendship and protest (these are three typical lyrics categories). The categorizer was developed by using the Weka libraries [Witten and Frank 1999]. In particular we employed the Naive Bayes algorithm included in Weka. In order to train the categorizer, we provided it with a corpus consisting of 60 songs manually divided into the three just mentioned categories. Once the categorizer was trained, it was able to automatically assign a list of pairs category/weight to a song, where the weight indicates the degree with which the song belongs to the category.

The annotation of pictures by the user is more complicated, although not that much as it seems to be. On the one hand, annotating the date of the picture, if it is taken with a digital camera, is an automatic processes, as the date is fixed by the camera itself. Thus, the user only has to annotate the pictures taken with a non-digital camera. On the other hand, annotating the topics of the pictures is something the user is starting to be accustomed to do, thanks to services like Flickr, Picasa, etc. where users tag pictures with keywords. Furthermore, there are many initiatives for the semantic annotation of pictures, mostly promoted by the flourishing semantic web. For instance, in [Schreiber et al. 2001] an approach and a related tool to make annotations in pictures is described. The pictures are annotated according to ontologies define in RDF Schema. Another example is presented in [Halaschek-Wiener et al. 2006], where a tool for the annotation and management of pictures in the semantic web is described. In particular, it facilitates the creation and publication of OWL annotations about the content of pictures in the semantic web.

Moreover, new techniques for automatically categorizing pictures according to the context where they appear (for instance, the text of the paragraphs around it) have appeared during the last few years. In this sense, a study about the functional categorization of pictures of web documents is described in [Hu and Bagga 2004].

On the other hand, there already exist prototypes of digital cameras that, apart from the purely visual aspects of the scene, allow the photographer to capture other aspects of the context such as the noise, temperature, movement, pollution, etc [Rost et al. 2005][Ljungblad et al. 2004][Hakansson et al. 2003]. These elements, which will be automatically annotated in future models of digital cameras, will permit new mechanisms of picture selection in Musimage.

In our case, we defined a simple ontology in order to annotate our personal pictures. Actually, our ontology is a taxonomy. Even though simple, a taxonomy allows us to perform powerful inferences by taking advantage of its hierarchical structure. Next we show an example. In our ontology the names of our friends are instances of the class ’friend’. In turn, the class ’friend’ is a descendant of the class ’loved one’, and it is, in turn, descendant of the class ’people’. The class ’loved one’ has other descendants (apart from ’friend’) such as ’relative’. Our pictures were manually annotated according to this ontology. In our picture collection several pictures were annotated with the tag ’Luis’ and the tag ’Luis’ is included as a friend in the ontology. The system takes advantage of the hierarchical structure of the taxonomy and is able to retrieve the pictures annotated with the tag ’Luis’ whether the system is looking for pictures of Luis, friends, loved ones or people (Luis and its ancestors in the taxonomy).

In order to illustrate the overall functioning of Musimage, next we show the ordered series of steps which are followed in a typical case of use:

1. The user selects a song or a list of songs to listen to.

2. The system analyzes the ID3 description of each song. The ID3 codifies meta-data in the mp3 files. These meta-data include the song title, artist, album, year, genre, comments, and they can even include the lyrics. The system obtains the four first mentioned meta-data.

3. The system retrieves the lyrics corresponding to each song from a lyrics server on the Internet. To find the lyrics, it uses the meta-data obtained from the ID3 description.

4. The system automatically assigns a list of weighted topics to a song by applying text categorization techniques to the song lyrics. The list of weighted topics is a list of pairs topic/weight where the weight indicates the degree with which the song belongs to this topic. To achieve the automatic categorization, the current system employs the Naive Bayes algorithm included in the Weka libraries [Witten and Frank 1999].

5. The system selects a set of pictures to be shown while each song is played, taking into account the following song characteristics: list of topics, year and temporal length.

6. The songs in the list are sequentially played, and for each song being played the corresponding selected set of pictures is displayed.

The next section details the process indicated in the 5th step, where the system selects the set of pictures to be shown while each song is played, taking into account several song characteristics (list of topics, year and temporal length).

4 Selection of pictures

Given a song s to be played, the system selects a subset of pictures to be shown while playing the song. In particular, n pictures are selected, where n is calculated as follows:

$$n = \frac{length(s)}{refreshTime}$$

where $length(s)$ is the duration of the song s and $refreshTime$ is the time that every picture is exposed. More specifically, the system selects the n pictures with the greatest value of similarity with respect to the song s. Similarity between a song s and a picture p is defined as:

$$similarity(s,p) = ((c_1(s,p) \times w_1) + (c_2(s,p) \times w_2))$$

where the similarity is calculated as the weighted addition of two different criteria, c_1 and c_2. c_1 is a criterion which measures the similarity by taking into account both the year when the song was recorded and the year when the picture was taken. On the other hand, c_2 is a criterion which measures the similarity by estimating the affinity between the topics of the song and the topics of the picture. Both criteria c_1 and c_2 are respectively weighted by w_1 and w_2, which are such that $w_1 + w_2 = 1$. These weights can be fixed by the user by employing the user interface. Next we show how c_1 and c_2 are calculated.

4.1 Year-based similarity

The criterion c_1 measures the similarity between a song s and a picture p taking into account both the year when the song was recorded and the year when the picture was taken. This similarity is defined as:

$$c_1(s,p) = \begin{cases} 0 & , \ d(s,p) < 0 \\ \frac{c-d(s,p)}{c} & , \ 0 \leq d(s,p) < c \\ 0 & , \ d(s,p) \geq c \end{cases}$$

where c is a constant value which defines a maximum number of years, such that if the picture p was taken c or more years later than the song was recorded, there is no similarity between p and s. The value of c can be fixed from the user interface. And $d(s,p)$ is defined as:

$$d(s,p) = year(p) - year(s)$$

where $year(p)$ is the year when the picture p was taken and $year(s)$ is the year when the song s was recorded.

Thus, by employing the criterion c_1, the similarity between a song and a picture is greater than zero if the picture was taken the same year that the

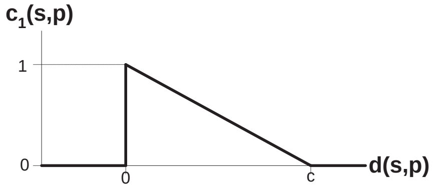

Figure 3: Year-based similarity

song was recorded or after that year, but less than c years later (see figure 3). Moreover, the nearer both dates are the greater the similarity is.

In order to define the just described function we informally interviewed 10 users. Thus, we think this kind of function is appropriate for using Musimage in a personal environment (for instance at home with a personal picture collection). Probably other functions could be defined for other contexts.

4.2 Topic-based similarity

The criterion c_2 measures the similarity between a song s and a picture p by estimating the affinity between the topics of the song and the topics of the picture. The song s is automatically categorised. As a result, for each category of songs, sc, a value $vcat(s, sc)$ is obtained, such that this value indicates the degree the song s belongs to the category sc. Let $\{sc_1, ..., sc_n\}$ be the set of categories of songs sc_i such that $vcat(s, sc_i) > 0$. On the other hand, each picture p is annotated with a set of categories of pictures. Let $\{pc_1, ..., pc_p\}$ be the set of categories of pictures that the picture p belongs to (i.e. the picture p is annotated with these categories).

The system utilises a table which stores the affinity between pairs (sc, pc), where sc is a category of songs and pc is a category of pictures. Let $affinity(sc, pc)$ be the value of the affinity between sc and pc stored in that table. This value indicates the degree with which we expect a picture of category pc to be triggered by a song of category sc. Thus, the topic-based similarity between the song s and the picture p is defined as:

$$c_2(s,p) = \sum_{i=1}^{n} \sum_{j=1}^{p} vcat(s, sc_i) \times affinity(sc_i, pc_j)$$

and values of $c_2(s,p)$ are later normalized between 0 to 1.

5 Applications and scenarios

As we mentioned before, we are exploring the indirection of information. We study how the user's actions can be translated, indirectly, into other actions

Figure 4: Musimage applied in a personal context

beyond the original user's intention. In this sense, Musimage translates, through the just described mechanisms, the original user's actions (selecting a list of songs to listen to) into new actions (showing a set of related pictures at the same time).

This kind of indirection of information can be applied to several different applications and scenarios. The next subsections present two different applications: enhancing spaces (the application which motivated this work) and transmission of presence (the application which was motivated by this work).

5.1 Enhancing spaces

Musimage was conceived as a system which augments the power of evocation of music by illustrating the user experience with pictures. Thus, Musimage first application is to enhance physical spaces (and therefore the user's experience) with new digital information.

This kind of application can be used in two kinds of scenarios. On the one hand, it can be employed in personal contexts like the user's home. In this case, Musimage accompanies the user in his recollection process by displaying pictures mainly from the user's personal picture collection according to the songs being played at the same time (see figure 4).

On the other hand, Musimage can be used in shared public spaces. For instance, it can be used in bars to augment the music played by the DJ with projection of pictures (see figure 5). In this case, Musimage illustrates the songs

Figure 5: Musimage applied in a shared public space (bar)

being played by displaying pictures from picture server on the Internet according to the songs. In this sense, Musimage could act as a provocateur of social relations and conversations.

5.2 Transmission of presence

While working on Musimage we realised that this kind of indirection of information could be used not only to enhance the user's space, but also to enhance the space of a remote user. In this sense, this approach would function as a social awareness system [Markopoulos et al. 2005]. Thus, we shifted towards the exploration of this kind of indirect information as a means of reinforcing the feeling of being accompanied. A first system, Emotinet [Ibanez et al. 2008], has already been developed as a natural extension of Musimage. In short, with a certain periodicity the user is presented, on a peripheral user interface (windows desktop or digital picture frame), with a new collage composed of pictures (see figure 6) indirectly triggered by their loved ones' actions. In particular the pictures are triggered by the songs they listen to and the text they write and read, while working on their PCs.

6 Conclusions and future work

In this paper we have described the design and development of Musimage, a novel system which displays pictures according to the songs being played at the same time. By using the interface the user selects the songs to be played, but the pictures are automatically selected. For each song to be played, the system selects a set of pictures, according to various criteria corresponding to some features of the song. In this sense, the pictures to be shown are, metaphorically,

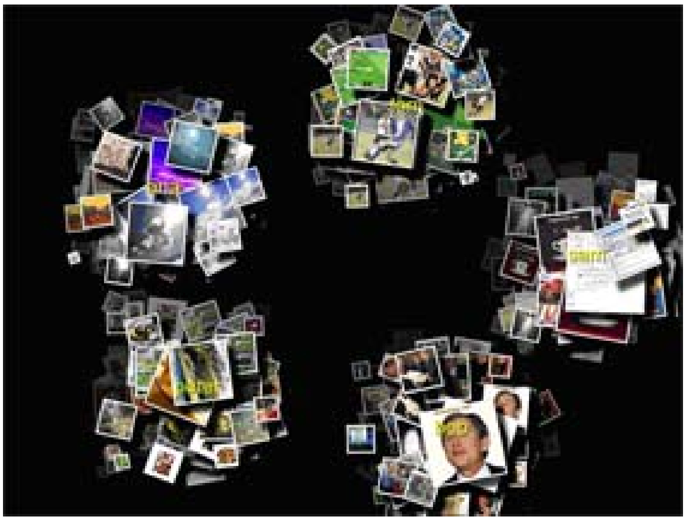

Figure 6: Collage composed of pictures indirectly triggered by the user's loved ones' actions. In particular the pictures are triggered by the songs they listen to and the text they write and read, while working on their PCs.

triggered by the songs. The system has already been developed and successfully tested. The system has been developed as a multi-agent system, which provides flexibility and extensibility.

The described system has already been employed successfully by the authors in individual contexts, and in the near future we plan to use it in shared public spaces. Our idea is to use Musimage in bars to augment the music played by the DJ with projection of pictures. We will explore it as provocateur of social relations and conversations.

As future work we will also investigate the use of other actions and media as input and output (apart from songs and pictures). Actually, in Emotinet not only the music played by the user is considered as input, but also the text the user writes and reads on his PC is considered as input as well. We will explore more abstract and dynamic visualizations as output. In this sense, we have recently proposed a novel approach to express emotions through the visualization of flock of virtual beings [Ibanez et al. 2007]. Emotions are represented in terms of the arousal and valence dimensions and they are visually expressed in a simple way through the behaviour and appearance of the individuals in the flock. In particular, the arousal value parameterizes the Reynolds' flocking algorithm, and the valence value determines the number of different colors in the flock. We carried out a study with users, whose interesting results are reported in [Ibanez et al. 2007]. We found that emotions can be expressed through the

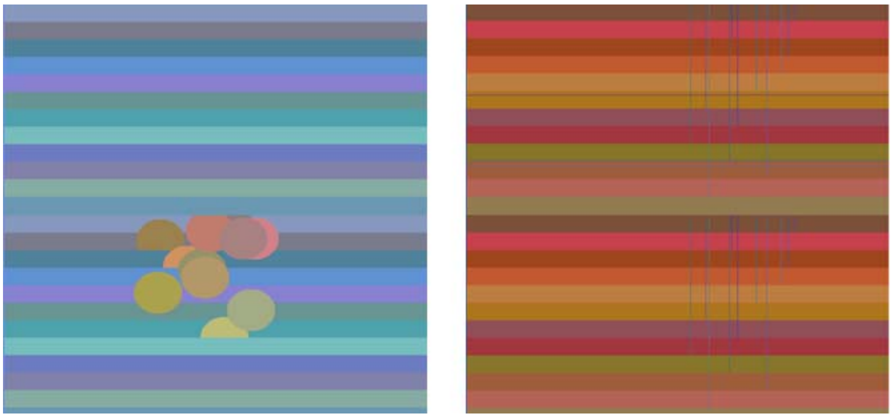

Figure 7: Artistic visualisation of emotions through the movement of digital entities and the number of different colors

movement of digital entities and the number of different colors. At the moment we are researching this kind of expression of emotions through the animation of more artistic entities (see figure 7).

We will also explore the use of the system for education purposes. Several subjects might be better learnt and taught by exploiting both audio and visual channels. Further research will be conducted in order to explore this perspective.

Acknowledgements

We would frankly like to thank Leticia Lipp for generously proofreading. We would also like to thank the reviewers for their wise comments. We think they have clearly improved the quality of the paper in several aspects. Furthermore, one of the reviewers suggested the application of the system for education purposes which is included in the future work.

References

[Bellifemine et al. 2007] Bellifemine, F.L., Caire, G., Greenwood, D.:“Developing multi-agent systems with Jade”; Wiley Series in Agent Technology, John Wiley and Sons Ltd, 2007.

[de Ruyter 2003] de Ruyter, B.:“365 days ambient intelligence research in Homelab”; Neroc Publishers, Eindhoven, the Netherlands, 2003.

[Dertouzos 2001] Dertouzos, M.L.:“The unfinished revolution: Making computers human-centric”; HarperCollins Publishers, 2001.

[Ferber 1999] Ferber, J.:“Multi-agent system: An introduction to distributed artificial intelligence”; Addison Wesley, 1999.

[Hakansson et al. 2003] Hakansson, M., Ljungblad, S., Holmquist, L. E.: "Capturing the invisible: Designing context-aware photography"; Proceedings of the 2003 conference on Designing for user experiences, DUX'03 (New York, NY, USA), ACM, 2003, 1-4.

[Halaschek-Wiener et al. 2006] Halaschek-Wiener, C., Golbeck, J., Schain, A., Grove, M., Parsia, B., Hendler, J.A.: "Annotation and provenance tracking in Semantic Web photo libraries"; IPAW (Luc Moreau and Ian T. Foster, eds.), Lecture Notes in Computer Science 4145, (2006), Springer, 82-89.

[Hu and Bagga 2004] Hu, J., Bagga, A.: "Categorizing images in Web documents"; IEEE MultiMedia 11, 1 (2004) 22-30.

[Ibanez et al. 2008] Ibanez, J., Serrano, O., Garcia, D.: "Emotinet: a Framework for the Development of Social Awareness Systems"; To appear in P. Markopoulos, W. MacKay, and B. de Ruyter (Eds.), Awareness Systems: Advances in Theory, Methodology and Design, Springer, Series on HCI.

[Ibanez et al. 2007] Ibanez, J., Delgado-Mata, C., Gomez-Caballero, F.: "A Novel Approach to Express Emotions through a Flock of Virtual Beings"; Proceedings of the 2007 International Conference on Cyberworlds (Hannover, Germany), IEEE Computer Society, October 2007, 241-248.

[LIAFLine 2004] "Liafline: Newsletter of the Long Island Alzheimer's Foundation", February 2004.

[Ljungblad et al. 2004] Ljungblad, S., Hakansson, M., Gaye, L., Holmquist, L. E.: "Context photography: Modifying the digital camera into a new creative tool"; Proceedings of CHI '04: CHI '04 extended abstracts on Human factors in computing systems (New York, NY, USA), ACM, 2004, 1191-1194.

[Markopoulos et al. 2005] Markopoulos, P., de Ruyter, B., Mackay, W.E.: "Awareness systems: Known results, theory, concepts and future challenges"; Proceedings of CHI '05: CHI '05 extended abstracts on Human factors in computing systems (New York, NY, USA), ACM Press, 2005, 2128-2129.

[Marzano and Aarts 2003] Marzano, S., Aarts, E. (eds.):"The new everyday: Views on ambient intelligence"; Uitgeverij 010 Publishers, 2003.

[Mitchell 2000] Mitchell, W.J.:"E-Topia"; The MIT Press, August 2000.

[Rost et al. 2005] Rost, M., Gaye, L., Hakansson, M., Ljungblad, S., Holmquist, L. E.: "Context photography on camera phones"; Proceedings of UbiComp 2005 (Tokyo, Japan), September 2005.

[Schreiber et al. 2001] Schreiber, A.T., Dubbeldam, B., Wielemaker, J., Wielinga, B.: "Ontology-based photo annotation"; IEEE Intelligent Systems 16, 3 (2001) 66-74.

[Schulkind et al. 1999] Schulkind, M.D., Hennis, L.K., Rubin, D.C.: "Music, emotion and autobiographical memory: They're playing your song"; Memory & Cognition 27 (1999), 948-955.

[Weiser 1991] Weiser, M.: "The Computer for the twenty-first century"; Scientific American 265, 3 (1991) 94-104.

[Weiser 1993] Weiser, M.: "Ubiquitous computing"; IEEE Computer 26, 10 (1993) 71-72.

[Witten and Frank 1999] Witten, I.H., Frank, E.:"Data mining: Practical machine learning tools and techniques with Java implementations"; The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, 1999.