Tobias Siebenlist, Kathrin Knautz

Chapter 15. The critical role of the cold-start problem and incentive systems in emotional Web 2.0 services

Abstract: Some content in multimedia resources can depict or evoke certain emotions in users. The aim of Emotional Information Retrieval (EmIR) is to identify knowledge about emotional-laden documents and to use these findings in a new kind of World Wide Web information service that allows users to search and browse by emotion. The cold-start problem arises in systems where users need to rate or tag documents when there is not sufficient information to draw interferences from these documents and the existing ratings or tags for those. The cold-start problem can be divided into three situations: cold-start system, cold-start user and cold-start item. In this paper we will refer to the first one and the last one. We will discuss the cold-start problem of the specialized search engine for emotional-laden multimedia documents called MEMOSE in this article and provide a solution using a combination of using content-based features to overcome the start-up phase and an incentive system for getting and keeping users motivated to use MEMOSE.

Keywords: Emotional information retrieval, indexing; retrieval, content-based information retrieval, concept-based information retrieval, cold-start, incentive system, gamification, Media EMOtion Search, MEMOSE

Tobias Siebenlist (corresponding author), Research Associate and PhD Student, Department of Information Science, Heinrich-Heine-University, tobias.siebenlist@uni-duesseldorf.de
Kathrin Knautz, Research Associate and PhD Student, Department of Information Science, Heinrich-Heine-University

Introduction

In the early days of the World Wide Web in the 1990's, only a few experts were able to make content available online. In 1996 there were about 250,000 sites at 45 million worldwide users. Although these sites were accessible online, their use was passive in nature because user-generated content was rather uncommon due to the complexity of the deployment. This changed with the emergence of new and easy-to-use services at the beginning of the 21st century. It was now possible for users to generate and publish content online by themselves. Via their PCs or laptop computers and additionally via mobile telephones, smart phones, netbooks or tablet computers users can nowadays produce content like videos, images and music at almost any place and time and share it with the online community.

Some content in multimedia resources can depict or evoke certain emotions in users. The aim of Emotional Information Retrieval (EmIR) and of our research is to identify knowledge about emotional-laden documents and to use these findings in a new kind of World Wide Web information service that allows users to search and browse by emotion. For these reasons, the special search engine MEMOSE was developed (Knautz, Siebenlist, & Stock, 2010; Knautz, Neal, Schmidt, Siebenlist, & Stock, 2011). MEMOSE is the first prototype which tries to index emotions in multimedia documents based on user input. Users select emotions of a controlled vocabulary containing ten basic emotions. The intensity of an emotion can be adjusted via slide controls. Studies show that users are able to index images (Schmidt & Stock, 2009) and videos (Knautz et al., 2010) consistently.

While users are able to index consistently, there exists a cold-start problem. If a new document is added to the system, how can it be found when it has not been rated sufficiently? Until a certain number of users have tagged the document emotionally, it is not possible to separate the appropriate emotions. An approach to solving this problem is presented here, which utilizes a combination of content-based and concept-based methods. Content-based information retrieval works with basic features which are given in the document (e.g., colour and shape of images). Additionally we take a look at implementing an incentive system (Roinestad et al., 2009; Wash & Rader, 2007) into MEMOSE in order to motivate the users to rate more documents.

1 MEMOSE – a specialized search engine for emotional-laden multimedia documents

MEMOSE is a search engine that focuses on multimedia documents (images, videos and music) and the emotional indexing of those documents (Knautz et al., 2010; Knautz et al., 2011). Content is added solely by the users of MEMOSE. Multimedia documents can be imported from supported web services via their Application Programming Interfaces (APIs) or by using an upload form for such documents. For retrieval purposes the documents need to be tagged. Either the tags are retrieved from the supported web service or the uploading user assigns tags while using the upload form. The specialty about MEMOSE is the focus on emotions. When a new multimedia document is added to the system it needs to be emotionally tagged. We use a controlled vocabulary of 10 basic emotions (sadness, anger, fear, disgust, shame, surprise, desire, happiness, fun and love) which are assigned using slide controls in a range from 0 to 10, with a value of 10 representing the highest intensity regarding the respective emotion. Further we distinguish between the emotion that is *expressed* (what is displayed) and *felt* (what the user feels while interacting with the document). This variety of choices leads to 20 slide controls, allowing the user to make accurate decisions about the emotions represented within a multimedia document. When searching for a document, the user can enter search terms and choose one or more emotions to search for. The search terms will be compared with the documents' tags, and the emotional values are compared using the average value from the users' ratings. As of today, the relevance ranking depends solely on the emotional intensity of the documents assigned by the users' emotional tagging. A document has to be emotionally tagged by a defined number of users in order to appear within the search results. When starting MEMOSE as a publicly available service, this leads to the problem that nearly no results can be found because the documents have not been rated by a sufficient number of users to obtain reasonable results.

Also, since MEMOSE depends on the documents that are contributed by the users, we need mechanics to overcome the start-up phase until it gets established with usable content. The problem here is that when starting the service, no documents exist in the database, which leads to an unusable service. When nothing can be found, there is no need for a search engine. We have provided some sample data that the developers and students have put into MEMOSE, but for satisfactory results, there is a need for real users and their content. Although we need users to contribute content, we also need them to emotionally tag the available documents. The users must be motivated to use MEMOSE regularly and to provide new content as well as tag documents with emotions.

2 The cold-start problem

When rolling out a new service to the public on the Web, we have to deal with the question: How can users be motivated and attracted to actively use the service and spread the word about it? Especially when the service builds upon user-generated content and the users' activities, the service needs active users in order to be successful. Starting a new service without having a sufficient amount of user-generated data to draw inferences for the content is known as the cold-start problem.

According to Park, Pennock, Madani, Good, and DeCoste (2006), three cold-start situations exist: cold-start system, cold-start user, and cold-start item. The first one describes the situation when a new service starts and there are only small numbers of users and documents. The cold-start user situation exists when a new user registers, and the service has no or little knowledge of the user. The third situation happens when a new document is added to the service and there is no more than one emotional rating by the user who added it. The first and the last one are the cold-start problems we have to deal with in MEMOSE. A cold-start user situation can be neglected in MEMOSE, because the retrieval is not based on information about the users but on the documents' emotional ratings given by the users. The cold-start system and the cold-start item situations will both be considered with the approach described in this paper.

When a new collaborative service appears on the Web, the number of users is usually small. The service must become known in order to get more users. This is even more true if the service relies on the activities of the users, as with MEMOSE. People need to add media documents and tag them in order to make MEMOSE more interesting and useful for other users. The cold-start problem describes the situation of missing knowledge about items (users, documents, etc.) which disables the service to make recommendations about these items. The following question arises: If a new document is added to the system, how can it be found when it has not been rated sufficiently? Until a certain number of users have tagged the document emotionally, it is not possible to separate the appropriate emotions. Thus no further inferences and classification can be drawn for the available documents by any means. The users need to become motivated to tag the documents and to actively use the service. Therefore, we provide a combination of content-based and concept-based methods. In this part of the paper, we will concentrate on content-based retrieval and the usage of extracted features to support the users' choices.

3 Content-based retrieval

The field of content-based retrieval (CBR) deals with the extraction of so-called "features" from multimedia documents. These features are characteristics of the documents and can be used to index, search and find those documents for example within a retrieval system. In contrast to concept-based retrieval, which uses additional information like tags, controlled terms, notations or other information in textual form, CBR is used for automatic indexing of the documents' resources. As this can be done automatically, no further intervention is needed. It is available for usage in an automated indexing process. Besides the automatic indexing of textual data, the field of multimedia information retrieval deals with the indexing of multimedia documents (images, videos, music tracks) using those characteristics. One application of CBR is to find multimedia documents by using similar examples as the query. A problem that has to be noted is the huge amount of data that has to be processed in order to retrieve features from multimedia documents. Thus, this is a complex process that takes time. The features that can be extracted from characteristics within the documents are called low-level features. MEMOSE supports three multimedia types: images, videos and music. Frequently used features for images are colour, texture and shape. For videos, scene and shots are used. For extracted frames from videos, the same features as mentioned with images can be used. From music, features like pitch, rhythm and harmony can be extracted.

In this paper, we will focus on content-based image retrieval (CBIR) as a beginning to the usage of content-based information in MEMOSE. With this decision, the following statements are focusing only on images. According to Knautz et al. (2011, p. 218), "the information which can be deduced from these low-level features is limited, and therefore only partly suitable for an analysis of the semantic and emotional content". Due to this limitation, we propose the use of content-based retrieval not as the single solution, but as a supportive element in the whole process of indexing and retrieval.

A common CBIR method is to create summary statistics about an image document. These statistics represent, for example, colour usage, texture composition, shape, structure, and further values. Regarding the features of an image, a distinction has to be made between global features and local features (Al-Khatib, Day, Ghafoor, & Berra, 1999). Global features contain an extracted value for a feature that represents the whole image with every pixel included. In comparison, local features represent not the whole image, but a smaller region that is part of the image.

3.1 Available features for content-based retrieval

The basic features that can be extracted from images are colour, texture, shape and spatial information. These low-level features are often used when working in the field of CBIR. We will give a short description of these features below, before we explain what features will be used in our approach. A possible presence of noise is a limitation of this work; the images are processed as they are.

3.1.1 Colour

The term “colour” describes light that comes from some kind of source, either directly or being reflected or transmitted. The human perception of colour is three-dimensional, and is modelled as a mixture of red, green and blue (RGB). Every pixel of an image has a value for colour; thus, colour is the property of a pixel. According to Jörgensen (2003), “a color histogram is a representation of the distribution of colors in the image”. A histogram is a graphical representation of the distribution of data. A colour histogram of an image is a visualization of the tonal distribution; it gives an estimation of the distribution of the colours in the image.

Colour is a very obvious and very important feature of an image. Additionally, the values for the colour of an image can be extracted without many computational costs. Multimedia documents with a similar content often have similar or identical colours (Jörgensen, 2003). For example, an ice-bear has about the same colour on every image, though the surrounding can differ. Narrowing down the setting even more to a fixed setting (e.g. an ice-bear with the sea in the background) leads to more similar histograms for different images. Colour histograms are described as one of the most basic features, and they are widely used in CBIR (Deselaers, Keysers, & Ney, 2008).

3.1.2 Texture

An image can be described as a composition of different textures. The textures are distributed in regions within the image. Regions contain some image features that can be extracted and used for indexing and retrieval purposes. Every texture describes a pattern in an image, depending on criteria like repetition, orientation and complexity. When extracting texture features from an image, the aim is to determine dominant components, whereas the colour feature could be extracted for every pixel of an image, a texture is a property of a region of pixels. Tamura et al. (1978) have determined through psychological studies that humans respond best to coarseness, contrast, and directionality. The so-called Tamura textures deal with these three features. They are computed separately and put together for example into a false-colour image where every feature is related to one colour of the RGB scale.

3.1.3 Shape

Shapes within images describe objects with a physical structure. With the use of shape features, different shapes can be compared by using the contour and region of the shape. To determine shapes in an image, segmentation or edge detection and analysis is frequently used. Due to light effects and masking of objects, there often is no possibility of determining the shape of the images' content reliably. Shapes can occur in many different forms. Global features are derived from the entire shape; for example, roundness or circularity. Local features are computed by processing the shape, and lead to information like the size of the shape or its corners. The information about the existence of a shape and the differentiation between shapes can be found in images by searching for changes to a signal's intensity. Pavlidis (1988) describes the following five situations, when a change in signal refers to a shape:
1. Change in reflectivity of a surface
2. Change in orientation of a surface
3. Combined changes in the relation of the object and the viewer
4. Direct occlusion of one object by another
5. Indirect occlusion (shadow) of one object by another

3.1.4 Spatial information

An image as a whole displays some kind of content. The features we have discussed so far dealt with the image as one big entity, even though it is split into shapes or different textures have been found. Spatial information can be used to identify different areas which carry different importance. For these areas, separate histograms can be computed, regarding the areas of interest in an image. A popular and quite promising idea to detect local areas is to compute points of interest. These are salient points in local areas of the image. The technique used to detect and encode local features in images is called scale-invariant feature transform (SIFT) developed by Lowe (1999, 2004). Thus the features are so-called SIFT-features. In order to compute those SIFT-features, candidate points need to be found, orientations need to be assigned, and local reference coordinate systems need to be built around them. The final SIFT-features will be extracted from this area around the candidate points, preserving information about this local reference. The features can be encoded in a scale, rotation and location invariant manner. According to Rüger (2010, p. 57), "a typical image exhibits in the order of 2000 key-points".

3.1.5 Face detection

On a higher level of feature extraction, faces on pictures can be extracted reliably in images using the open source library OpenCV (Open Source Computer Vision). Whenever an image contains faces that at least partially show the front of the face, the faces can be recognized, and the appropriate areas in the image are returned as results. Having the coordinates of the face, this area can be extracted and further investigated. Within the face, the eyes can be recognized easily as prominent points. Several approaches to automatic emotion recognition have used facial expression as a basis to gather information about the person's emotions (Black & Yacoob, 1995; Essa, 1997; Mase, 1991; Tian, Kanade & Cohn, 2000; Yacoob & Davis, 1994). Azcarate, Hageloh, van de Sande, and Valenti (2005) describe a method for recognizing the facial emotion of a person by processing video content and extracting motion units to compare them to a fixed list of sample motion units. Chia (2010) uses images as well as a combination of support vector machines and multinomial logistic regression to compare and evaluate relevant extracted features. Generally, the positions and states of face elements like eyes, eyebrows, and lips and so on are extracted and compared to a sample image that prototypically expresses an emotion. According to Chia (2010, p. 1), "[f]acial expression recognition is concerned with the recognition of certain facial movements without attempts to determine or presume about the underlying emotional state of the agent". This approach does not cover the problem that facial expressions may result from physical exertion and that there is the need for more than just visual data to analyse an emotion. But in a system like MEMOSE, this is the only data we can build upon, so we will neglect this fact here.

Both attempts use the Cohn-Kanade database (Kanade, Cohn & Tian, 2000) as dataset. This dataset consists of 500 image sequences from 100 subjects. Accompanying meta-data include annotation of FACS (Facial Action Coding System) action units and emotion-specified expressions. This database can be obtained for research and non-commercial use for free.

3.2 Deciding which features to use

Deselears et al. (2008) have done an experimental comparison of features for image retrieval where they give an overview of many image features, including the ones presented here. The performance evaluation of features with five different benchmark image databases for CBIR showed that the colour histogram performed very well for all colour tasks (Deselaers et al., 2008, p. 15) According to Deselaers et al. (2008, p. 15), "color histograms … clearly are a reasonably good baseline for general color photographs." Some approaches (mainly SIFT-based ones) have outperformed the colour histogram approach. The use of these approaches requires much higher computational costs, whereas the colour histogram needs only minimum computational costs. Because of these great results for the approach using colour histograms and because of the low computational costs, we decided to use this approach in MEMOSE. The low computational costs allow the service to compute the histogram immediately after an image document has been uploaded so that this image can be used for recommendations right away without waiting for an image processing period. Another interesting approach is the usage of SIFT-features. Due to higher computational costs and more complex image processing tasks, SIFT-features have not yet been implemented in MEMOSE. As we do not restrict the content of the documents to a specific domain, we can only rarely make use of face detection algorithms. Nevertheless, all images are automatically examined with a regularly repeating task. If faces are found, this information is stored in the database with coordinates where the faces occur in the image. Because there is a certain amount of false positives, the users have the possibility of reporting images when faces have been recognized falsely.

3.3 Finding emotions in low-level features

Many multimedia documents depict or evoke emotions in the observer. Because MEMOSE is focused on emotions, we will discuss the context of emotions within low-level features shortly. We have already shown that there are numerous low-level features that can be extracted automatically from multimedia documents at higher or lower computational cost. Higher-level features (e.g. objects, names or even emotions) are much more difficult to extract automatically because when the observer looks at or listens to a multimedia document, there is some kind of understanding and connecting with the knowledge that has been gathered so far. This difficulty is known as the semantic gap (Smeulders, Worring, Santini, Gupta, & Jain, 2000; Zhao & Grosky, 2001). To bridge this gap is one of the biggest and most difficulty challenges in multimedia retrieval. According to Wang and Wang (2005, p. 1), "emotion semantics is the most abstract and highest, which is usually described in adjective form". Wang and Wang further explain that emotions are closely related to cognitive models, culture backgrounds and aesthetic standards.

The interpretation of extracted low-level features like colours, shapes and textures is based on their implication in art (Arnheim, 1983; Itten, 1973). A typical approach to using knowledge from this field is to implement these connections as algorithms and programs that obtain the ability to recognize and categorize the low-level features appropriate. For example, “horizontal lines always associate with static horizon and communicate calmness and relaxation” (Wang, Yu, & Zhang, 2004, p. 6407). Wang and Wang (2005) mention four issues in emotion-based semantic image retrieval: Performing feature extraction, determining users’ emotions in relation to images, creating a usable model of emotions, and allowing users to personalize the model. These issues are widely noted in the literature regarding content-based work: “defining rules that capture visual meaning can be difficult” (Colombo, Del Bimbo, & Pietro, 1999, p. 52). Further, Dellandrea, Liu, and Chen (2010) mention three issues in classifying emotion present in images: “emotion representation, image features used to represent emotions and classification schemes designed to handle the distinctive characteristics of emotions” (para. 1). The different methods that have been developed in the field of content-based emotional image retrieval included mechanisms like Support Vector Machines (SVM) (Wang et al., 2003; Wang et al., 2004), machine learning (Feng et al., 2010), fuzzy rules (Kim et al., 2005), neural networks (Kim et al., 2007; Kim et al., 2009; Dellandrea et al., 2010), and evidence theory (Dellandrea et al., 2010). While the majority of these studies suggest promisingly effective results, the systems have been very minimally tested by users before they are reported in research. Additionally, personal preferences and individual differences are not incorporated into the work (Knautz et al., 2011). As previously mentioned, our content-based features approach concentrates on the colour features and thus on colour histograms because the extraction of these does not need high computational cost and has a low complexity. Thus the extracted histograms can directly after computation be used in the service without the need to wait for processing with more elaborate methods.

As the users tag documents with emotions, we can start using learning algorithms that are trained with the users’ tagging behaviours and thus could lead to better recommendations and improved relevance ranking. This needs to be elaborated within a study with real users and real data so that the learning can be founded on a sufficient number of users that have tagged the documents with emotions.

4 Similarity measures

Since we have decided to use the colour histogram approach in order to get features from images, the next step is to think about a similarity measure for comparison of the images. According to Jörgensen (2003), the most common similarity metrics for evaluating colour similarity are histogram intersection and weighted distance between colour histograms. But before we concentrate on such elaborate methods, we need to start with rather basic similarity measures. The Minkowski or L_p distance is a metric on the Euclidean space and defined as follows:

Let $x, y \in R^n$:

$$L_p(x, y) = \sqrt[p]{\sum_{i=1}^{d} |x_i - y_i|^p}$$

Special cases of this norm are L1 known as the *Manhattan norm* and L2 known as the *Euclidean norm*. These two norms are in general widely used to measure similarity between objects.

Each dimension or image feature is independent of each other and of equal importance. Thus, every bin of the colour histogram is treated independently, and does not account for the fact that certain pairs of bins correspond to features which are perceptually more similar than others. (Suhasini et al., 2009). Nevertheless, the Minskowski distance and their respective special cases are used in various fields of applications, not only the comparison of images or their histograms.

Another common measure in information retrieval is the cosine similarity, defined as follows:

$$d_{cos}(v, w) = \frac{v * w}{L_2(v) L_2(w)}$$

The cosine similarity is a measure of the similarity between vectors. The cosine of the angle between these vectors determines the similarity between them. This kind of measure is often used in text-mining and for text statistics as well as data-mining.

The next similarity measure is a special similarity measure for histograms. The partial histogram intersection is defined as follows:

$$d_{phi}(v, w) = \frac{\sum_i \min(v_i, w_i)}{\max(L_1(v), L_1(w))}$$

Here, v and w are not necessarily normalized histograms, but they should at least be in the same range to yield usable results. The components of v and w are expected to be non-negative (Rüger, 2010). Deselaers et al. (2008) follow Puzicha et al. (1999) by comparing histograms using the Jeffrey-divergence (JD) or Jensen-Shannon divergence (JSD), which is defined as follows:

$$d_{JSD}(H,H') = \sum_{m=1}^{M} H_m \log \frac{2H_m}{H_m + H'_m} + H'_m \log \frac{2H'_m}{H'_m + H_m}$$

where H and H' are the histograms to be compared and H_m is the m_{th} bin of H. The Jeffrey-divergence (JD) is the symmetric and stable version of the Kullback-Leibler divergence (KL). The Kullback-Leibler divergence is regarded as a measure of the extent to which two probability density functions agree. The KL has been suggested by Ojala et al. (1996) as an image dissimilarity measure, and measures how inefficient on average it would be to code one histogram using the other as the true distribution for coding.

Choosing the right similarity measure for the right task is not obvious. The results of the presented measures need to be examined within MEMOSE after an evaluation period where users are faced with recommendations based on the different measures. Since Deselaers et al. (2008) obtained very good results using the Jenson-Shannon divergence, we took this similarity measure as our first choice. The results for recommendations and relevance ranking using this similarity measures seem promising, but due to the limited scope of the colour histogram comparison, the recommendations as well as the added documents to the list of results are not always understandable and reasonable to the users.

5 Using content-based features to solve the cold-start problem

5.1 Extraction and storage of the content-based features

Since we decided to use the colour histograms as the main concept-based feature in MEMOSE, we need to extract it from the image documents. This is achieved by computing the colour histogram of every image that is present in the database. These RGB values are saved in the database in order to have them accessible when needed. Right after that, a background task is started in which the colour histogram is compared to every other colour histogram in the database. As we have decided to use one of the presented measures for measuring similarity between two colour histograms, we can easily adopt the stored values to the respective formula. For the purpose of accessing this information in many search queries, this data is also permanently stored in the database, because this information will not change. As the database grows and the computing of similarity values for each of the existing image documents takes longer, this task is outsourced into a background task that runs after inserting a new image. The user does not get any notice of this process. After the process has finished, the newly added image and its colour histogram are available for being used within the service. The histogram for each image in the database of MEMOSE is created using Python and the Python Imaging Library (PIL). All the histograms are normalized. The histogram of each image consists of an array of integers that describe the colour values in an RGB scale.

Another feature we decided to use is the face recognition using OpenCV. Every document that contains one or more faces is marked in the database, and the coordinates for every recognized face are stored. For these recognized faces, we try to detect facial expressions and to match them to emotions. Until now we have not used the Cohn-Kanade database. The use of this database and further algorithms for the detection of facial emotion expression will be implemented later.

5.2 Using content-based features for relevance ranking

The goal of this task is not to identify emotion-specific histograms, but to find similar images to those that already have been emotionally tagged, and to let the user decide if there is some similarity in those images. As we have described before, the tagged documents will be within the results only if a certain number of users have emotionally tagged these. The problem that arises is: how can these documents get tagged by a sufficient number of users when the documents cannot be found? The solution here is divided into two parts. First, the new documents which have not yet been rated by a sufficient number of users are presented in a special area of the website marked as "new documents" or "tag these documents". Apart from the documents that have been recently added to the system, the users get an overview of all documents that they have not tagged yet and that have not been tagged by a sufficient number of users to enter the list of results. The second part uses the content-based features described earlier. Having the images within the list of results give them a better chance to attract attention and to get more users to tag them. When the user starts a query, the result is computed regarding the search terms and the chosen emotions. Before the ranked list of results is returned to the user interface, it will be extended with some documents that are similar regarding the colour histogram similarity to the retrieved results. As mentioned before, the results of the colour histogram for every document compared with the other documents' colour histogram are stored in the database so that a query can be processed very fast. A script controls the addition of similar images. Only documents that provide a certain similarity to the retrieved documents for the query are added to the list of results. The amount of similarity can be adjusted by a variable parameter. A good value needs to be found by evaluation. If no document reaches this similarity border, a small number of documents are also added to the results. These documents are also the ones with the highest similarity value below the parameter. Thus the results contain in each case at least some documents that need to be tagged by more users to become available as a search result. Figure 1 visualizes this process.

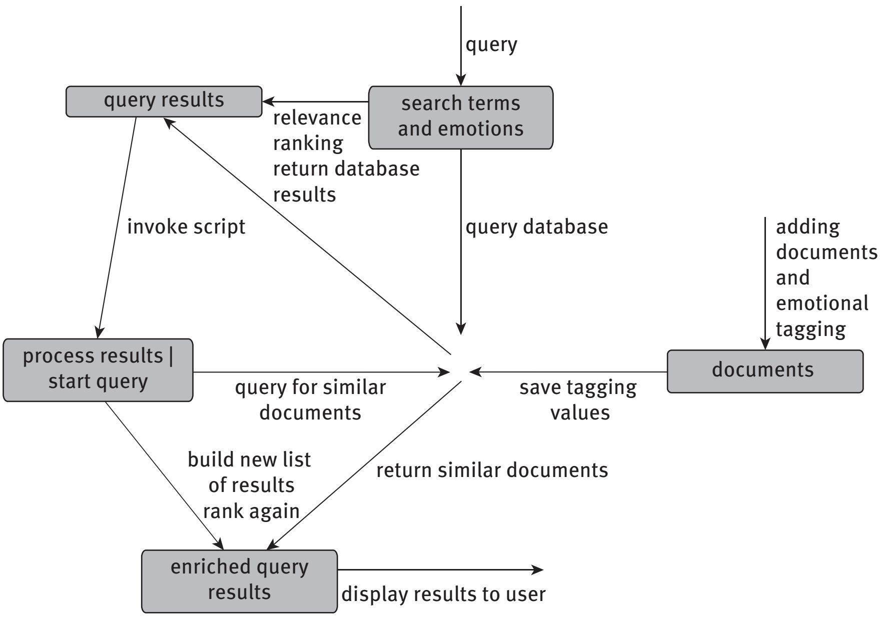

Figure 1. Relevance ranking with addition of similar documents.

As long as a document is not tagged frequently enough by users, there will be no information about the emotional tagging. Instead of a scale and the arithmetic mean value for the selected emotions, there will be just a message that asks the user to tag this document. The message's text is adapted to the queried emotions; for example, the following message could be displayed instead of a scale: "Does this image depict or evoke love? Tell us what you think and tag it!" As the comparison of documents is based on colour histograms, the documents with a high similarity value have a high probability of displaying similar content. Thus, it makes sense to ask the user if the same emotions are appropriate for the proposed document. If there are no or too few documents to compare a document to, further investigation of the colour histograms is needed. Therefore, the colour distribution of the image can be divided into warm and cold colours, and these can be applied to a group of emotions.

5.3 Content-based recommendations

In addition to the use of comparing colour histograms for the purpose of relevance ranking, this data can be used to make recommendations to the users. The recommendations will be shown to the users aside from the list of results for a query. As the queries of the users are logged, a search profile of every user is built. Using the information in this profile, the user will get recommendations regarding the search behaviour and the emotional tagging that has been done so far by the user.

A comparison between the tagging behaviour of the users can be drawn in order to receive documents that one user has already tagged, but another similar user has not. These documents will be recommended to users that have not tagged it yet. In order to solve the cold-start item situation, documents with too few ratings are recommended for tagging. For these recommendations, a comparison of the users' tagging behaviour and the image histogram of the document are utilized.

5.4 Concluding remarks about the use of content-based features

So far, we only have begun integrating content-based retrieval and comparison of colour histogram into MEMOSE. First tests with users show promising results. The recommendations have been correct in the majority of cases. But due to the small number of participants and documents, no reliable conclusion can be drawn from these early tests. An extensive evaluation of these new features is in preparation, and is necessary to obtain conclusive results. Deselaers et al. (2008) showed that a comparison using colour histogram does lead to usable results, although it should not be used solely as method of giving recommendations.

We have shown an approach for using content-based features to solve the cold-start problem. We have to deal with two situations with MEMOSE: cold-start system and cold-start item. Both situations can be addressed using the techniques described here. The recognition of faces within images using OpenCV provides, in most cases, good results. The users are informed that faces have been recognized on an image and can report false positives. However, the automatic emotion recognition using facial expressions is a hard task. We will do further research, since we have a sufficient number of images that contain one or more faces using the Cohn-Kanade database. The extraction of face elements does work, though. Therefore, we propose using the analysis of images with colour histograms as a supportive technique within a broader solution that combines this histogram comparison with an incentive system that will now be explained further.

6 Incentive system

MEMOSE is designed to enable users to find videos, music and pictures, which convey or evoke certain emotions. In order to facilitate this emotional search, the media files have to be attributed with emotional properties. This is achieved by providing a slider-based emotional tagging tool to users. As a direct consequence of this approach, a high degree of user activity becomes a necessity. This part of the paper shows a multi-layered approach to an incentive system, incorporating concepts of role-playing games and gaming platforms with the purpose of motivating users to become active parts of the community and provide the necessary emotional tagging.

6.1 Depending on active users

When the World Wide Web had just seen the light of day in the early 1990s, very few people could provide online content. While these websites could be accessed online, their usage was of a purely passive nature, because user-generated content was too complex to be realized at that time. However, this changed in the beginning of the 21^{st} century with the creation of new services that are much easier to use. Using such services, users are now able to actively generate content, instead of simply consuming it. The role of the user thus changed from a passive content consumer to that of an active content provider. With the use of smartphones, computers and netbooks, users of today's age can produce content in the form of videos, images and music at nearly any time and place and distribute it through the online communities of social media services. Similar to other Web 2.0 platforms, there is also a reciprocal, interdependent relationship between MEMOSE and its users. MEMOSE requires active users who provide content and thus create an attractive user interface.

On the other hand, users also desire the platform to be a source of multimedia documents or a stage on which they can present their own works. Unfortunately, this dependency in Web 2.0 is usually unbalanced. Because a user can choose among a huge variety of different services, the creator of a platform is forced to try to convince the user of the singularity of his service and provide incentives for its use. This imbalance is not necessarily a bad thing, since it forces the provider to continually improve the platform and thus make it more attractive for the users. Active user participation is especially important for MEMOSE, because the emotional search of multimedia documents is largely based on user-generated metadata. Only when new images, videos or music pieces are indexed with a sufficient number of emotional ratings can the underlying concept of the platform unfold.

6.2 Gaming concepts in an information context

By indexing multimedia files emotionally, users provide a service that cannot be automated to that extent. Besides the previously described extraction of low-level features, we thus require an intellectual indexation. By using a pre-determined set of ten emotions, the emotional indexation by the user follows the categorization model. This is supplemented with the possibility of judging the intensity of an emotion through the use of the dimension model. We differentiate her between the emotions depicted by the medium and those that a user experiences while looking at it or listening to it (Knautz et al., 2010; Knautz et al., 2011). MEMOSE not only enables a user to search for media and these two different types of emotions, but also allows him to find out more about the emotional expressiveness of his own media by sharing them with the community. The system thus provides an incentive for users to integrate their own contents into the system. However, the tagging of other users' contents carries no similar incentives. Since the indexing of content follows a collaborative approach in MEMOSE, we have thus to create incentives for users.

In the last few years, game mechanics from digital roleplaying games implemented in online services have become a popular method to increase user participation. Well-known examples are locating games like foursquare.com, a service in which users transmit their current location through the GPS-locating function of their smartphones. Status messages about the restaurants they have visited or the distances they have covered turn elements of everyday life into a game. Besides the supplementary analysis features, users are also rewarded with points and badges for sharing with the community. The goal of such services is to inexpensively learn the habits and likes and dislikes of a new generation of users in order to be able to bind these users to a product or brand.

Measures like these are often summed up with the controversially discussed term *gamification*. Critiques (e.g. Bogost, 2011; Robertson, 2011) begin with the neologism *gamification* itself. For them, the term is a semantic turgidity, since most applications only offer a point system in conjunction with achievements. According to Robertson (2011), "[g]ames give their players meaningful choices that meaningfully impact on the world of the game" (para. 6). Real game mechanics, desirable goals and rule systems are absent, for which reason Robertson prefers the term *pointification*.

Gamification thought leader Gabe Zichermann, on the other hand, defines gamification as "[t]he use of game thinking and game mechanics to engage users and solve problems" (Zichermann & Cunningham, 2011, p. XII). According to this definition, gamification does mean turning elements of everyday life into a game, but rather to use aspects from games successfully in other areas:

> Gamification makes it possible for big brands and start-ups alike to engage us in meaningful and interesting ways, with an eye on aligning our personal motivations with their business objectives. The net product of this effort will be more engagement, better products and – generally – more fun and play in all spheres of our lives. (p. XII)

On our platform, we thus also use game mechanics and give users incentives to participate in emotional tagging. It is important to consider after all that "[p] eople play not because they are personally interested in solving an instance of a computational problem but because they wish to be entertained" (von Ahn & Dabbish, 2008, p. 60).

7 Engage users – multi-layered incentive system

In their paper "Designing games with a purpose", Ahn and Dabbish (2008) characterize three motivational factors which seem to argue for the implementation of game mechanics for data generation. They mention that despite the growing existence of Internet resources, humans can often solve problems which are (as of yet) impossible for computers to solve. A third argument is that people like to play games, and spend a lot of time doing it at the computer (von Ahn & Dabbish, 2008). Taken together, these three factors open up the possibility of letting the multimedia documents in MEMOSE be indexed emotionally by networked individuals.

This leaves the question of how we can successfully use the mechanics of gamification to motivate users to participate and interact. It should be noted that we differentiate between the intrinsic and the extrinsic motivation of a person

(Barbuto & Scholl, 1998). If a person is intrinsically motivated, it means that they are doing something that satisfies their own interests, something that they enjoy doing. The act is thus autotelic. If, on the other hand, a person is extrinsically motivated, they are doing something because they expect to gain certain advantages or rewards for it. Krystle Jiang (2011) summarizes the relation between gamification and motivation: “More and more companies are converting to game-based marketing tools and are met with great success. The reason is because game mechanics are strong extrinsic psychological motivators in a domain where there is little to no intrinsic motivation” (p. 2).

We can thus say that it is important to use extrinsic motivators in order to address intrinsic needs. The company Bunchball, which specializes in gamification, explains which game mechanics can be used to satisfy specific desires of users (Table 1). It is easy to see that all game mechanics (e.g. points) address one primary game dynamic (reward), but they also each cover different motivations (status, achievement, competition, altruism).

Game Mechanics	**Human Desires**					
	Reward	**Status**	**Achievement**	**Self-Expression**	**Competition**	**Altruism**
Points	x	•	•		•	•
Levels		x	•		•	
Challenges	•	•	x	•	•	•
Virtual Goods	•	•	•	x	•	
Leader boards		•	•		x	•
Gifting & Charity		•	•		•	x

Table 1. Interaction of basic human desires and game play. The “x” signifies the primary desire a particular game mechanic fulfils, and the “•” indicates the other areas that are affected (Bunchball, 2010, p. 9).

In a prototype version of the incentive system of MEMOSE, we integrated points, levels and leader boards in order to cover as many motivations as possible with few mechanics. Below, we will show how the implementation works.

7.1 Points: Hatching a monster, rewarded by experience points

Points form an important, if not a vital, part of incentive systems. Points and the rewards connected to them have the function of motivating users to interact in the ways the operator desires. This form of game mechanics (*token economy*) has its origin in behaviour therapy and is based on the principles of operative conditioning (e.g. Ayllon & Azrin, 1968; Kazdin, 1977). Points are positive, generated multipliers, which are used to develop and maintain a certain behaviour. By awarding points for every successfully produced output of a user, they become an immediate motivational game mechanic. Points are rewards and provide feedback about past performance. Furthermore, they provide a way for users to make a name for them and distinguish themselves from others by increasing their points.

Depending on the requirements according to which points are awarded, different kinds of point can be identified. Zichermann and Cunningham (2011) differentiate between five different kinds of points: experience points (XP), redeemable (RP), skill, karma and reputation. XP in particular play an important role. They are awarded for every successful activity and inform the players as well as the operator about rank, status and abilities. Redeemable points, on the other hand, work as a sort of currency that users can save up and subsequently spend on certain things. The third kind of points, skill points, can be earned via specific activities in the system and are closely linked to XP and RP. They function as bonus points which are awarded for meta-successes. According to Zichermann and Cunningham, karma points are used only rarely. They can be earned through evaluations of contributions and can be spent on polls for system improvements. The last and most complex kind of point that Zichermann and Cunningham (2011) mention is reputation points. Integrity and consistency are very important to make reputation points work. They are awarded for receiving positive evaluations of some form from other users (e.g. a positive critique of the user's pictures).

As a first step, we integrated XP into MEMOSE. After his registration a user is prompted to choose a monster egg. This egg basically contains a personal MEMOSE pet, and is linked directly to the user's account (Figure 2). The user can now cause the egg to hatch by accumulating enough experience points. XP can be earned by tagging media with emotions; thus, we have developed a first incentive to use the system actively. Once the monster has hatched, further XP allow it to develop and reach higher evolutionary stages. Thus, a monster's development stage becomes the indicator of the user's activity and a sort of status symbol within the MEMOSE community (Figure 3). Amongst other things, newly uploaded media by users whose monster has reached a higher stage of development are given priority placement in the *New Media* view. As a result, copious tagging of other people's media increases the probability that one's own media will be emotionally tagged.

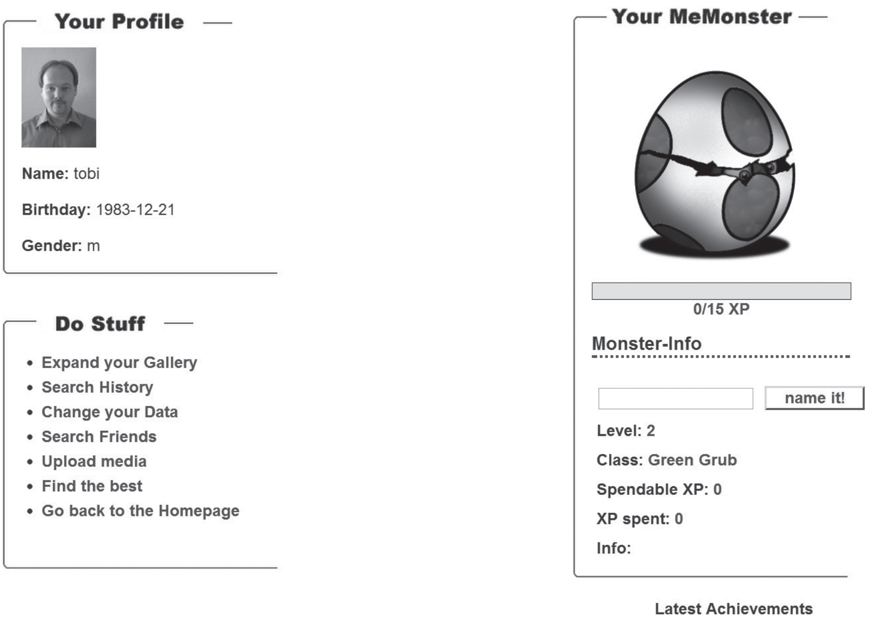

Figure 2. User profile with MeMonster.

Furthermore, experience points can be used as a form of currency on the platform, e.g. for buying a better placement of one's image in the *New Media* view (not to be confused with the placement in the search results). Redeemable points should be introduced for such transactions, but in order to avoid unnecessary complexity in the system, we have decided to restrict ourselves to XP in the early stages. This policy confronts users with the choice between improving their placement in the short term through an investment of XP, or developing their monster further in order to achieve a better placement in the long term. The integration of experience points and their use as digital currency is one of the cornerstones of the MEMOSE incentive system, and provides lots of room for its extension.

7.2 Levels: Noob or pro, show your progress

According to a theory by Latham and Locke (1990), human acts are goal-oriented. In their goal setting theory, they assume that conscious behaviour is steered by individual goals and serves a purpose. Challenges increase levels of performance and lead to satisfaction upon success. This satisfaction influences the level of commitment a person feels, which then initiates new actions.

Their research has shown that, once established, it makes no difference whether a goal was assigned (from someone else) or self-initiated (Latham & Locke, 1990; Klein, Wesson, Hollenbeck, & Alge, 1999). What is important, though, is to provide feedback about the performance level with respect to the goals, in order to increase the task-specific self-efficacy (Bandura, 1986). People with a higher level of self-efficacy set higher goals for themselves and have a higher level of commitment (Latham & Locke, 2002).

These research results can be applied to the development of incentive systems. The accumulation of points can serve as positive reinforcement to motivate users to initiate further interactions. The implementation of levels that can be reached serves a sign of progress. The pursuit of higher levels through the collection of points is based on the basic principle of goal-setting theory, which states that human acts are goal-oriented. It should be noted that higher goals are harder to reach and require a more complex solutions (Latham & Locke, 1990; Zichermann & Cunningham, 2011).

The possibility of the MEMOSE monster reaching various development stages reflects the platform's level design. Each higher level increases the number of new points necessary to attain said level. The levels themselves function as a progress marker that reflects a user's performance (Figure 2 and 3). The level of performance indicated the user's personal status within the community. Thus users are always anxious to improve said status. At this time, we have realized six evolutionary stages of the MeMonster, which will be further extended in the future. A progress bar enables the user as well as visitors of his profile page to see the current evolutionary stage and point total (Zichermann & Cunningham, 2011). The progress bar also shows how many points are needed to rise to the next level. It hopefully incites the user to earn the necessary points and discover the next evolutionary stage of his personal avatar.

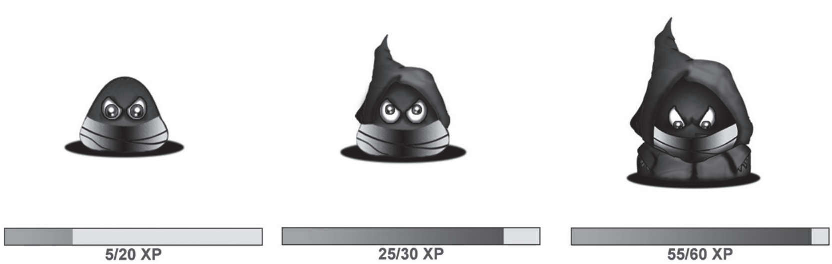

Figure 3. MEMOSE monster in its first evolutionary stages (hatch by accumulating enough experience points).

7.3 Achievements: Rewards for XP and levels

A further possibility for motivating users is to allow them to unlock achievements. Achievements (or badges) reward the user for specific activities or reaching certain milestones. Like the MEMOSE monster, achievements also originated from video games and have been used by large gaming platforms like Xbox LIVE, PlayStation Network, or Steam for some time now to motivate users (Jakobsson, 2011). Similar to points, achievements act as positive reinforcement. However, they motivate those users who have already unlocked an achievement more than those who have not. It is thus important to award the first achievement for reaching a simple goal (e.g. registering). Zichermann and Cunningham (2011) list the following reasons for the success of *achievements*:

> In addition to signalling status, people want badges for all kinds of reasons. For many people, collecting is a powerful drive. Other players enjoy the sudden rush of surprise or pleasure when an unexpected badge shows up in a gamified system. A well-designed, visually valuable badge can also be compelling for purely aesthetic reasons. (p. 39)

The statement can be supplemented with the summary of Antin and Churchill (2011), who explain five basic functions of badges:
- **Goal-Setting**: human behaviour is goal-oriented, users like to collect points and unlock badges in order to reach their goals;
- **Instruction**: lists of achievements inform the user about the possibilities of using the system;
- **Reputation**: the number and nature of badges allows others to draw conclusions about the experience, competency and trustworthiness of the user;
- **Status**: badges represent past successes and are accessible to others without necessitating verbal interaction; difficult achievements can lead to a better status;
- **Group Identification**: attaining difficult achievements can cause a feeling of solidarity between users. (pp. 2–3)

These days, many online services use achievements to meticulously protocol when users reach various milestones. The user thus profits from a sort of gallery to showcase his achievements and possibly gain recognition (Medler, 2009). While collecting achievements satisfies a human desire, analysing the unlocked achievements allows the service operator to gain valuable information. Various user groups can be differentiated, depending on which achievements were unlocked. Jakobsson (2011) differentiates between casuals, hunters and completists. Zichermann and Cunningham (2011) identify explorers, achievers, socializers and killers. Identifying various user types facilitates adjusting game mechanics and thus also satisfying and reinforcing user needs. Because of these reasons, we have similarly implemented achievements in MEMOSE. If, for example, a user reaches the list of top taggers, he earns a corresponding achievement with the title *Tag it like it's hot*. Even if the user fails to stay in the list, his achievement has been eternalized in his achievement gallery (Figure 4). The achievements are further linked to experience points, rewarding users with a certain number of points for each achievement. Like the concept of experience points, this system is easily extendable and should help to motivate users long-term.

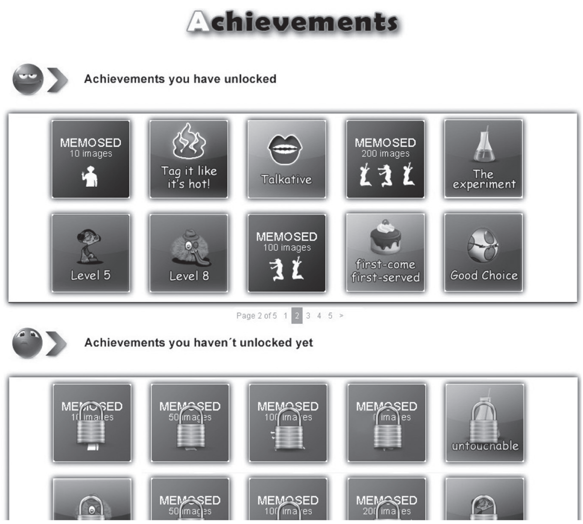

Figure 4. Achievement systems.

7.4 Leader boards: Social rewarding for users

According to Locke and Latham (1990), users can be motivated through challenging and specifically defined goals. The implementation of leader boards as a game mechanic is another way to formulate goals and motivate users (Hoisl et al., 2007; von Ahn & Dabbish, 2008; Zichermann & Cunningham, 2011). Furthermore, ranking lists inform the service provider that a game is being played, and how it is used (Zichermann & Linder, 2010). Ranking lists compare different achievements of users, and are usually easily understood. Competing with others and trying to reach first place can satisfy different human desires like status, grandstanding or competitiveness. Hoisl, Aigner, and Miksch (2007) point out, however, that while this mechanic can be used to stimulate motivation extrinsically, intrinsic motivation is a necessary prerequisite. The ranking lists implemented in MEMOSE cover different areas: most active tagger, most active uploader, and so on.

We are planning another kind of ranking list, as proposed by Zichermann and Cunningham (2011). They deem the comparison with direct friends in an online service as more interesting for users than the comparison with everyone. People tend to compete with other people around them. This is true for the digital world as well. Furthermore, such a personal ranking list is very motivating, since one's own name is always listed. Leader boards which apply to the whole user base may promise greater reputation, but they can also have a demotivating effect on people who hold a very low rank.

7.5 Comments: Expressing thoughts

Since the applications of emotionally indexed multimedia documents are manifold, it is apparent that very different user groups will exist. For example, emotional aspects play a vital part in music therapy. Its fields of application range from the therapy of psychotic patients with clinical depression (Lai, 1999; Hsu & Lai, 2004) or geriatric psychiatric patients (Lee, Chan, & Mok, 2010; Clair & Memmott, 2008; Short, 1995) to the rehabilitating treatment of neurological illnesses and manualized treatments of tumours (Burns, 2001).

Another important application of emotionally indexed multimedia documents can be found in the advertising industry. Targeted and conscious influencing of people causes conscious or subconscious desires. In this case, factual product information fades into the background. Methods of suggestion (McChesney, 2008) call the subconscious mind. However, in this first instance, MEMOSE was developed for social media users.

Recent studies (Bischoff, Firan, Nejdl, & Paiu, 2008, 2010) show that if we assign categories like time, place or type to tags, we find a large discrepancy between assigned tags and searched tags – particularly in the area of opinions/qualities. Thus, 24 % of Flickr queries consist of affective tags, but only 7 % of the documents have been indexed with affective tags (Bischoff, Firan, Nejdl, & Paiu, 2010). Users are searching for emotionally charged content, but they cannot find it. MEMOSE shall try to bridge this gap. It shall also allow users to exchange opinions and comments, as they do on other platforms (Flickr, YouTube).

We consider MEMOSE to be a part of Web 2.0 and a website "which encourage[s] user-generated content in the form of text, video, and photo postings along with comments" (Cormode & Krishnamurthy, 2008, p. 4). Registered users can add a comment to pictures of their own choosing. Unregistered visitors can access these contents, but cannot add comments. In this way, we pre-emptively reduce the number of undesirable comments and spam messages, while reducing user-unfriendly mechanisms like CAPTCHAs.

For logged in users, a form under each image offers a quick means of commenting. Each comment is cleaned of malicious code and hyperlinks before being linked to its author and the image in the database. The presentation of the comment in the browser is managed via AJAX, thus eliminating the need of reloading the page. The published comment also contains the publishing date as well as a link to the author's profile. In future versions, an avatar will accompany the nickname. Comments are ranked by date in a descending manner, thus allowing visitors to quickly trace the history of user reactions to a picture.

The comment system is only a first step towards connecting users with each other. Further measures towards strengthening the community shall follow (Kim, 2000; Zichermann & Linder, 2010; Zichermann & Cunningham, 2011). It is one of many functions of MEMOSE to encourage users "to create an account in order to more fully engage with the site" (Cormode & Krishnamurthy, 2008, p. 11).

8 Concluding remarks

Carrying game mechanics over from the gaming world into other economic and social areas is becoming increasingly common. The term used to describe this phenomenon, *gamification*, is being discussed controversially. The main point of criticism is the lack of rules and concrete goals. Nonetheless, implementing gaming elements in online services proves very successful at motivating users to generate data. These data are then used by companies and serviced to bind users to a certain brand, product or service. We believe that MEMOSE has the potential to be used by interested users even without such gamification elements. Studies show that emotional contents are searched for, but not found, in Web 2.0 services like YouTube, Flickr and Last.fm. A need for emotionally indexed media does thus exist. It is furthermore in the nature of many people to express their originality. By offering the possibility of uploading one's own media, MEMOSE creates a way for users to act out their creativity and individuality.

However, MEMOSE is still in the early stages of its development and depends on collaborative user input (e.g. indexing). Active user participation is important, because the emotional search is largely based on user-generated metadata. The extraction of low-level features has more of a support role. In order to overcome the beginning difficulties of a social media platform and motivate users to participate, we thus decided to make use of the implementation of game mechanics. By rewarding users and using positive reinforcement we hope to encourage the "desired" behaviour: the emotional tagging of media. Specifically, we have implemented points, levels, achievements, leader boards and comments in MEMOSE. Furthermore, we have added "Share on Facebook" buttons. We are planning to further enhance the social, community-oriented aspects by integrating a friendship system, sensitive high score lists and competitions for uploading media with emotional content. All of these measures are based on the idea that human behaviour has a purpose and is steered by individual goals and desires. These goals can be image cultivation, reputation or recognizing complex patterns, amongst others. Extrinsic methods (e.g. collecting points and achievements) are linked to intrinsic desires (e.g. the wish to complete something). We are aware that the game mechanics presented here are only a start and require extension. Nevertheless, game mechanics offer us a means to motivate users and generate data.

Acknowledgements

We wish to thank Daniel Guschauski, Daniel Miskovic, and Jens Terliesner, whose valuable support has helped us with working on MEMOSE.

References

Al-Khatib, W., Day, Y. F., Ghafoor, A., & Berra, P. B. (1999). Semantic modeling and knowledge representation in multimedia databases. *IEEE Transactions on Knowledge and Data Engineering, 11*(1), 64–80.

Antin, J., & Churchill, E. (2011, May 7–12). Badges in social media: A social psychological perspective [Blog post]. Retrieved from http://gamification-research.org/chi2011/papers

Arnheim, R. (1983). *The power of the center: A study of composition in the visual arts*. Berkeley, CA: University of California Press.

Ayllon, T., & Azrin, N. (1968). *The token economy: A motivational system for therapy and rehabilitation.* New York: Appleton-Century-Crofts.

Azcarate, A., Hageloh, F., van de Sande, K., Valenti, R. (2005). *Automatic facial emotion recognition*. Retrieved from http://staff.science.uva.nl/~rvalenti/projects/mmis/Automatic%20Facial%20Emotion%20Recognition.pdf

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* New York: Prentice Hall.

Barbuto, J., & Scholl, R. (1998). Motivation sources inventory: Development and validation of new scales to measure an integrative taxonomy of motivation. *Psychological Reports, 82*, 1011–1022.

Bischoff, K., Firan, C., Nejdl, W., & Paiu, R. (2008). Can all tags be used for search? *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (pp. 203–212).

Bischoff, K., Firan, C., Nejdl, W., & Paiu, R. (2010). Bridging the gap between tagging and querying vocabularies: Analyses and applications for enhancing multimedia IR. *Journal of Web Semantics*, *8*(2), 97–109.

Bunchball. (2010, October). Gamification 101: An introduction to the use of game dynamics to influence behavior. [Blog post]. Retrieved from http://www.bunchball.com/gamification/gamification101.pdf

Burns, D. (2001). The effect of the bonny method of guided imagery and music on the mood and life quality of cancer patients. *Journal of Music Therapy*, *38*(1), 51–65.

Chia, J. (2010). *Emotion recognition through facial expressions.* Retrieved from http://www.cs.ubc.ca/~johnchia/emotion-recognition.pdf

Clair, A., & Memmott, J. (2008). *Therapeutic uses of music with older adults.* Silver Spring, MD: American Music Therapy Association.

Colombo, C., Del Bimbo, A., & Pietro, P. (1999). Semantics in visual information retrieval. *IEEE MultiMedia*, *6*, 38–53.

Cormode, G., & Krishnamurthy, B. (2008). *Key differences between Web 1.0 and Web 2.0*. Retrieved from http://www2.research.att.com/~bala/papers/web1v2.pdf

Dellandrea, E., Liu, N., & Chen, L. (2010). Classification of affective semantics in images based on discrete and dimensional models of emotions. *Proceedings of the IEEE International Workshop on Content-Based Multimedia Indexing (CBMI)*.

Deselaers, T., Keysers, D. & Ney, H. (2008). Features for image retrieval: An experimental comparison. *Information Retrieval, 11*(2), 77–107.

Feng, H., Lesot, M. J., & Detyniecki, M. (2010). Using association rules to discover color-emotion relationships based on social tagging. *Proceedings of Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 544–553).

Hoisl, B., Aigner, W., & Miksch, S. (2007). Social rewarding in wiki systems – motivating the community. *Proceedings of the 2nd International Conference on Online Communities and Social Computing* (pp. 362–371).

Hsu, W., & Lai, H. (2004). Effects of music on major depression in psychiatric inpatients. *Archives of Psychiatric Nursing, 18*(5), 193–199.

Itten, J. (1973). *The art of color: The subjective experience and objective rationale of color*. New York: Van Nostrand Reinhold.

Jakobsson, M. (2011). The achievement machine: Understanding Xbox 360 achievements in gaming practices. *Game Studies. International Journal of Computer Game Research, 11*(1). Retrieved from http://gamestudies.org/1101/articles/jakobsson

Jaing, K. (2011). *The dangers of gamification: Why we shouldn't build a game layer on top of the world*. Retrieved from http://krystlejiang.files.wordpress.com/2011/07/the-dangers-of-gamification.pdf

Jörgensen, C. (2003). *Image retrieval: Theory and research*. Lanham, MD: Scarecrow Press.

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)* (pp. 46–53).

Kazdin, A. (1977). *The token economy.* New York: Plenum Press.

Kim, A. J. (2000). *Community building on the Web: Secret strategies for successful online communities.* Boston: Addison-Wesley Longman Publishing Co.

Kim, E. Y., Kim, S., Koo, H., Jeong, K., & Kim, J. (2005). Emotion-based textile indexing using colors and texture. *Proceedings of Fuzzy Systems and Knowledge Discovery* (pp. 1077–1080).

Kim, N. Y., Shin, Y., & Kim, E. Y. (2007) Emotion-based textile indexing using neural networks. *Proceedings of the 12th International Conference on Human-Computer Interaction: Intelligent Multimodal Interaction Environments* (pp. 349–357).

Kim, Y., Shin, Y., Kim, S., Kim, E. Y., & Shin, H. (2009). EBIR: Emotion-based image retrieval. *Digest of Technical Papers, International Conference on Consumer Electronics*.

Klein, H., Wesson, M., Hollenbeck, J., & Alge, B. (1999). Goal commitment and the goal setting process: Conceptual clarification and empirical synthesis. *Journal of Applied Psychology, 64*, 885–896.

Knautz, K., Neal, D., Schmidt, S., Siebenlist, T., & Stock, W. (2011). Finding emotional-laden resources on the World Wide Web. *Information, 2*(1), 217–246.

Knautz, K., Siebenlist, T., & Stock, W. (2010). MEMOSE. Search engine for emotions in multimedia documents. *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 791–792).

Lai, Y. (1999). Effects of music listening on depressed women in Taiwan. *Issues in Mental Health Nursing, 20*(3), 229–246.

Lee, Y., Chan, M., & Mok, E. (2010). Effectiveness of music intervention on the quality of life of older people. *Journal of Advanced Nursing , 66*(12), 2677–2687.

Locke, E., & Latham, G. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision* (pp. 1150–1157).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110.

McChesney, R. W. (2008). *The political economy of media: Enduring issues, emerging dilemmas*. New York: Monthly Review Press.

Medler, B. (2009). Generations of game analytics, achievements and high scores. *Eludamos. Journal for Computer Game Culture, 3*(2), 177–194.

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based feature distributions. *Pattern Recognition, 29*(1), 51–59.

Park, S.-T., Pennock, D., Madani, O., Good, N. & DeCoste, D. (2006). Naive filterbots for robust cold-start recommendations. *Proceedings of the 12th ACM SIGKDD Iinternational Conference on Knowledge Discovery and Data Mining* (pp. 699–705).

Pavlidis, T. (1988). Image analysis. *Annual Review of Computer Science, 3*, 121–146.

Puzicha, J., Rubner, Y., Tomasi, C., & Buhmann, J. (1999). Empirical evaluation of dissimilarity measures for color and texture. *Proceedings of the International Conference on Computer Vision* (pp. 1165–1173).

Robertson, M. (2010, October 6). Can't play, won't play [Blog post]. Retrieved from http://www.hideandseek.net/2010/10/06/cant-play-wont-play/

Roinestad, H., Burgoon, J., Markines, B., & Menczer, F. (2009). Incentives for social annotation. *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia* (pp. 327–328).

Rüger, S. (2010). *Multimedia information retrieval*. San Rafael, CA: Morgan & Claypool Publishers.

Short, A. (1995). Insight-oriented music therapy with elderly residents. *The Australian Journal of Music Therapy, 6*, 4–18.

Smuelders, A.W.M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349–1380.
Suhasini, P.S., Krishna, K.S., & Krishn, I.V.N. (2009). CBIR color histogram processing. *Journal of Theoretical and Applied Information Technology, 6*(1), 116–122.
Tamura, H., Mori, S., & Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics, 6,* 460–472.
von Ahn, L., & Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM, 51*(8), 58–67.
Wang, S., & Wang, X. (2005). Emotion semantics image retrieval: An [sic] brief overview. *Proceedings of the 1st International Conference* on *Affective Computing and Intelligent Interaction* (pp. 490–497).
Wang, S., Chen, E., Wang, X., & Zhang, Z. (2003). Research and implementation of a content-based emotional image retrieval model. *Proceedings of the Second International Conference on Active Media Technology* (pp. 293–302).
Wang, W.-N., Yu, Y.-L., & Zhang, J.-C. (2004). Image emotional classification: Static vs. dynamic. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (pp. 6407–6411).
Wash, R., & Rader, E. (2007). Public bookmarks and private benefits: An analysis of incentives in social computing. *Proceedings of the Annual Meeting of the American Society for Information Science and Technology, 44*.
Zhao R., & Grosky, W. I. (2002). Bridging the semantic gap in image retrieval. In T. Shih (Ed.), *Distributed multimedia databases: Techniques & applications* (pp. 14–36). Hershey, PA: Idea Group Publishing.
Zichermann, G. (2011, April 26). The purpose of gamification [Blog post]. Retrieved from http://radar.oreilly.com/2011/04/gamification-purpose-marketing.html
Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in Web and mobile apps.* Sebastopol, CA: O'Reilly Media.
Zichermann, G., & Linder, J. (2010). *Game-based marketing: Inspire customer loyalty through rewards, challenges, and contests.* Hoboken, NJ: John Wiley & Sons.