



# MARIE CURIE – PROJECT MIESON

## *Periodic Report – Period P1*

This document provides a summary of the work carried out by the fellow, José San Pedro Wandelmer, during the outgoing phase of Project MIESON (grant agreement n. 243570) between the months of February 2011 and March 2012 (Months 1 to 14).

José San Pedro Wandelmer – Telefonica I+D

15/05/12

## PUBLISHABLE SUMMARY

Multimedia-focused online community websites, such as YouTube or Flickr, collect vast amounts of new data generated by users on a daily basis. User generated data consists of two complementary types of information: *content*, which comprises multimedia resources being shared by the community (e.g. videos in YouTube, photographs in Flickr), and *context*, which comprises metadata (e.g. tags, descriptions) and feedback (e.g. ratings, comments).

Project MIESON proposes a research framework focused on developing methodologies for the effective combination of *content* and *context-based features* in order to improve our understanding of multimedia content and provide enhanced methods to access it. Users generate high-level information and offer semantic cues about multimedia data that are difficult to obtain using content analysis methods only. By harnessing contextual data generated collaboratively in on-line communities, individual contributions are merged into a socially agreed set of descriptions and judgments. Data mined from collaborative websites can be used to create more complete feature representations of multimedia elements, fusing content and context-based features.

MIESON defines a sample application relevant to the multimedia research community as a case study for the proposed methodology: aesthetics modeling. *Aesthetics modeling* aims at investigating methods to determine the aesthetic value of multimedia content, photographs in particular. This application suits the proposed methodology in several orthogonal aspects. On the one hand, predicting aesthetic quality from purely visual information is a challenging task to perform automatically due to the *semantic gap*. The inclusion of higher-level contextual information into the statistical models is expected to significantly improve the accuracy of automatic learning methods in predicting such value. On the other hand, aesthetic assessment is a highly subjective task. By analyzing large-scale collections of community feedback online we can model an average, even democratic, view of what aesthetically pleasant content looks like.

Information access and discovery applications may take advantage of such aesthetic models to improve user experience and effectiveness. This could be achieved by leveraging information about image aesthetic quality to promote attractive content in image search scenarios, as a factor complementing query relevance.

MIESON proposes a user-centric research approach, focused on leveraging technology to enhance information retrieval and user experience. The objective is to increase the effectiveness of tools to access, discover or create multimedia content in real world applications. To this end, during the returning phase of MIESON the project will consider the development an application prototype that takes advantage of the proposed methodology to the ends previously mentioned.

So far the project has covered the technological objectives set for the outgoing phase. Methods to learn statistical models from the combination of content and context information have been studied for the proposed scenario of modeling aesthetics. Emphasis has been placed in carrying out training activities in the topics of statistical data analysis, machine learning, and multimedia content analysis. A significant number of communication activities, including publications in scientific journals and conferences, talks and

participation in panels have been carried out to disseminate the results generated within the scope of the project in the outgoing phase. Additionally, the fellow has attended a number of courses and tutorials aiming obtaining the skills required to complete the project objectives.

Three important results have been derived from this work. First, a methodological approach has been proposed that leverages unstructured contextual information from user generated content (specifically, user comments) for revealing domain knowledge. The proposed method is application independent. We have used it to learn about aspects that influence human aesthetics perception. Second, a large-scale evaluation of the performance of content and context-based features for modeling the aesthetic value of photographs, which reveals a clear dominance of contextual information over content in terms of classification accuracy. This study also considers a unified content/context representation that outperforms both of the aforementioned approaches. Third, a method to leverage aesthetic models to enhance image search results by incorporating aesthetics in the ranking model, which effectively extends *relevance* to improve search user experience as demonstrated by a user study.

Two scientific articles have been published so far, a journal article in ACM Transactions on Information Systems, a journal of reference in the Information Retrieval field with one of the highest impact factors in its field, and a full conference paper at ACM International World Wide Web Conference, WWW'11. Two additional articles have been submitted to Conferences and have been accepted for publication during spring 2012 (ACM International World Wide Web Conference, WWW'12 and ACM Conference on Human Factors in Computing, CHI'12). All these conferences are top-tier scientific events and are extremely competitive in terms of acceptance rate.

Regarding the training objectives of the fellowship, the fellow has attended 6 talks by prestigious members of the fields related to MIESON (Multimedia information retrieval, cloud computing, HCI), as well as 5 tutorials and 4 courses that have provided him with the research skills defined in his career development plan. In addition to publications, the fellow has interacted with several research groups at the outgoing host institution, including the CiteSeer group led by Prof. Lee Giles and the DTRA group led by Prof. Frank Ritter, mentored 2 students from Prof. James Wang's group, gave several talks at both host institutions and was an invited panelist at the IST Symposium 2011. Also, the fellow attended several scientific research events, including ACM Conference on Human Factors in Computing Systems and ACM Conference on Multimedia, that provided him with networking opportunities.

During the returning phase of the project, the fellow will focus on the practical aspects of the methodology researched. This will involve using knowledge gained during the outgoing phase and transferring it into a working mobile prototype application aiming at enhancing both the user experience and the effectiveness in information access tasks. In addition to this main scientific objective, the fellow will carry out dissemination activities in relevant international scientific events related to the results generated within the scope of the project. The fellow will also work on transferring the technology researched during the outgoing phase into the European host institution, Telefonica I+D. Given the market reach of Telefonica I+D, such tech-transfer activities are likely to have an impact worldwide on the way people access information.

## PROJECT OBJECTIVES FOR THE PERIOD

The outgoing phase of Project MIESON proposed the development of a novel methodology for the effective combination of *content* and *context-based features* to improve our understanding of multimedia information. We proposed the specific use case of characterizing the aesthetic value of multimedia content in order to develop more effective retrieval techniques and enhance user experience when accessing information. In particular, it aimed at developing automatic tools for predicting the attractiveness of visual content, and using such aesthetic measures to extend the notion of *relevance*. These aesthetics prediction models could be used to provide alternative ways to retrieve content by re-ranking, taking aesthetic features into account to promote attractive content in the result list. The work proposed included the collection of comprehensive data sets that would allow us to train and evaluate the technology developed. The goal was to leverage community knowledge and resources to create extensive test collections required for evaluating the analysis methodology proposed.

The objectives for the outgoing phase were divided into four work packages:

- **Work package 1 (Data Representation Schemes)** focused on studying the representation of multimedia content using the proposed combined approach: content and context-based features. Regarding content-based features, the following three aspects of multimedia content representation were identified as relevant: visual bag-of-words schemes, visual similarity-oriented features and metrics, and attractiveness features. Regarding context-based features, the project considered mining the large amount of user generated available from Web 2.0 sites, including: tags, geo-tags, user comments, descriptions, etc. Social aspects were also considered. An expected outcome of this WP was a document that defines a series of *content and context-based features* able to convey sufficient information to allow for machine learning techniques to perform better at understanding and classifying multimedia content (Deliverable **D1.1**).
- **Work package 2 (Data collection)** had the goal of gathering a sufficiently large collection of raw data from the Web to use in the development and evaluation stages (Work packages 3 and 4). The results formalized in D1.1 would be used to define the required data to be collected. Once the requirements for the collection were determined, a crawling of the required websites would be performed. The data collection process, including information about source where the data was retrieved from, tools used to get the data (e.g. API, languages, etc.), data fields gathered for each element (e.g. raw content, title, description, tags, etc.) and social aspects collected (e.g. category, group, owner, etc.), would be released as deliverable **D2.1**.
- **Work Package 3 (Learning from data)** dealt with the study of machine learning and statistical methods to exploit data collected in WP2. Multiple approaches for the effective hybrid representation of multimedia content were to be analyzed to optimize learning results. This WP would focus on studying the specific application scenario of *aesthetics modeling*. The aesthetic appeal of multimedia content is a subjective concept and, hence, the task of building aesthetic models could greatly benefit from the use of collaboratively created training sets: individual judgments of images would be merged generating a socially averaged score of visual appealing.

- **Work Package 4 (Evaluation)** considered the evaluation of all the developed technology. The evaluation would be performed using different techniques:
  - o *Quantitative approach:* evaluation approach based on quantitative accuracy metrics. Ground truth would be obtained from both community feedback available online as well as assessments provided by recruited participants.
  - o *User evaluation:* Accuracy metrics reveal many aspects of how methods perform. However, when dealing with user-oriented functionality such as results ranking, user-provided feedback is a valuable and relevant resource to have. User experiments would be designed where feedback from a set of selected participants will be used to assess the qualitative value of results achieved by the different methods proposed.

## WORK PROGRESS AND ACHIEVEMENTS

### 1 - Summary of Progress Towards Project Objectives

The outgoing phase of MIESON project took place in The College of Information Science and Technology at The Pennsylvania State University. During this phase, the fellow focused on investigating the data analysis methods defined in the project proposal. In this regard, the outgoing phase of the project has been a major success, covering all the work proposed in work packages 1, 2, 3 and 4.

In particular, for WP1 the fellow focused on studying the representation of multimedia content using the multi-modal approach proposed. He counted with the expertise of the outgoing host group, and was trained on the analysis of multimedia information for the extraction of relevant low-level visual content features. In addition to features related to content representation, the fellow also focused on studying metadata of relevance for the purposes of the project, completing the tasks established for this work package. Work package 2, dealt with the collection of datasets and was dependent on the results obtained in WP1. A detailed report of the current outcomes of these two work packages can be found in deliverables D1.1 and D2.1.

Work packages 1, 3 and 4 have been the most active during the outgoing phase. They encompassed the study of machine learning and statistical methods to analyze the data collected for WP2 using suitable feature representations from those studied in WP1. In addition to the extensive training received in these specific skills, the fellow conducted a series of studies focused on the *aesthetic modeling* problem setting. This work has produced very good results, including two article submissions to international conferences, and has covered the totality of objectives regarding aesthetics in the project, including:

- Leveraging unstructured contextual information from a social website (in particular user comments) to reveal community knowledge about aesthetics. This work follows

- the methodology proposed in the project to analyze user-generated content to better understand multimedia information. In particular, the most relevant factors influencing human aesthetic perception are identified. It is described in Section 1.1.
- Modeling aesthetics from multimodal image feature representations, combining content and context features. This work tackles the main analytical objective of WP3. It is described in Section 1.2.
  - Enhancing image search results by incorporating aesthetics in the ranking model. This work tackles another fundamental research question of the project: can we extend *relevance* using orthogonal aspects, such as *aesthetic value*, to improve search user experience? This work presents a real world application of aesthetic modeling and achieves significant gains for users. It is described in Section 1.3.

The fellow has generated additional scientific results relevant to the project objectives as part of ongoing work he had been carrying out prior to the start of MIESON. There are two specific works that are in consonance with the methodology and objectives proposed:

- Combination of content and metadata mining methods for propagating tags between adjacent multimedia elements in the network of resources that results from such analysis. This method is used to carry out automatic video tagging by spreading the wisdom of the community across the network of media elements.
- Combination of visual and text features for enhancing search results using multimodal queries. This work is in consonance with the goal of the project of taking advantage of multimodal representations, of the query in this case, for enhancing information access.

These two works generated two publications during the first year of the project, and are also briefly described in Section 1.4.

In summary, the outgoing phase has been very productive towards achieving the project objectives. The main work packages involved in this first phase were:

- Work Package 1: successfully completed and its output, D1.1, has been finished.
- Work Package 2: it has generated a new dataset and as well as deliverable D2.1. Licensing and copyright reasons prevent the dataset collected so far from being released. We discuss this matter in a subsequent section of this report.
- Work Package 3 & 4: the fellow has successfully completed the research objectives related to *aesthetic modeling* (described in Sections 1.1, 1.2 and 1.3) and has also dedicated some effort to *automatic tagging* (Section 1.4).

The rest of this section describes in more depth the most relevant works carried out by the fellow during the first year of this fellowship.

## 1.1 - Understanding Multimedia Information through the Analysis of Users' Comments in the Social Web

During work package 1, the fellow conducted an initial study of user feedback in one popular social website about photography: DPChallenge (<http://www.dpchallenge.com>). This preliminary study revealed that user comments convey highly valuable qualitative information about technical aspects of the photographs, many of them related to features relevant to their *aesthetic quality*. If properly leveraged, these comments are intuitively a very appropriate contextual feature to be used in the context of MIESON. They comprise human judgments, possibly a more accurate information source about aesthetic value than visual or other contextual features.

Given the potential of this source of contextual information, the fellow investigated the analysis of users' comments in online communities as a way to extract information about multimedia content. The objective was two-fold:

1. Getting hands-on experience with *technologies and methods to analyze unstructured textual* information. Social participation on the web has been increasingly gaining in popularity as the Web has started incorporating efficient tools to allow for such community feedback. Deliverable D1.1 includes an extensive description of the many forms of feedback that are enabled by current social-oriented websites. Unstructured textual information, such as user comments, is a popular feedback mechanism and it *is widely available* in the social web.
2. Investigating the viability of leveraging user provided unstructured feedback as an efficient information source for content representation and multimedia understanding. In contrast to structured user feedback (e.g. ratings or tags), the analysis of unstructured user feedback is challenging given the unrestricted nature of the contents provided by users. Therefore, leveraging this source of information is more challenging and may present difficulties in its application to machine learning.

We established the goal of developing a method to *reveal the main factors that influence image aesthetic perception* by analyzing user discussions about photography available online. With this analysis, the fellow aimed at revealing a commonly agreed set of photography aspects that users identify as relevant for their judgment of visual aesthetic value.

### *Methodology*

For each photograph in DPChallenge, the following forms of community feedback are accessible: a numeric rating (average of all ratings provided by individual users) and text comments written about the photograph. Photography ratings serve as proxies for aesthetic quality of photographs. Text comments, on the other hand, can be leveraged to infer user judgments about the best and worst aesthetics aspects of photographs.

This latter step, inferring user judgments, requires transforming the unstructured user comments into structured elements easier to analyze automatically. To this end, the fellow carried out an extensive literature review of natural language processing (NLP) tools and, in particular, sentiment analysis and opinion mining methods. Sentiment analysis methods are



intuitively suitable for the proposed problem given their ability to reveal both the aspects being discussed (e.g. color, composition, saturation) and the orientation of the opinions about these aspects or factors (e.g. nice, poor, beautiful). The fellow developed a python-based sentiment analysis tool extending on a previous state-of-the-art method judged as especially suited to the considered problem setting during the literature review process.

The goal of sentiment analysis tools is to transform comments from a set of English sentences into a list of duplets of the form (*factor*, *opinion*) that can be easily analyzed using standard statistical and machine learning methods. The sentiment analysis algorithm implemented uses hidden Markov models (HMM) to infer the function of each word in the sentence, therefore identifying factors (words that describe specific characteristics of the item being commented, such as color, composition or lighting) and opinions (either positively or negatively- oriented).

The final strategy implemented combined this sentiment-based representation of photographs with a supervised learning algorithm to predict photograph ratings from comments. This prediction model was then used to compute factor relevance based on their relative contribution to the prediction. This approach enables gaining important domain knowledge and better understanding of the concept being studied, photography aesthetics in this case.

To achieve the desired result, the fellow reviewed existing literature on factor analysis and aspect rating. These two analytical approaches look at discovering what are the main factors (or aspects) that users value more about a specific product or service. This process is normally performed by carefully designing an experiment where users are asked to choose between different instances of the same product, each instance having different values for each analyzed factor. The main difficulty in our problem setting was that, while normal factor analysis deals with a relatively small number of factors (no more than 5), our preliminary study of the discussion in DPChallenge revealed over 20 different aesthetic factors being actively commented by the community. The number of combinations needed to avoid any confounding effects was not feasible to implement as a user experiment.

The fellow resolved this scalability issue by considering an automatic approach based on statistical analysis. This approach considers the factors commented for each photograph in DPChallenge (e.g. composition, color, sharpness, tone, etc.) as independent variables, which take values depending on the orientation of the opinions of commenters. For example, a photograph with many users commenting positively on the *composition* but negatively on *tones* would have a positive score for the *composition* variable and a negative score for *tone*. Using these variables,  $x_i$ , we can create a model to predict photograph ratings from comments. Multilinear regression is used to this end:

$$Y = \beta_0 + \beta_1 x_i + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

Image rating,  $Y$ , is the dependent variable in this model, generated as a linear combination of aesthetic factors. In this expression,  $\beta_j$  denotes the regression coefficients to be estimated and  $\epsilon$  the prediction error, which is assumed to be normally distributed. The values  $\beta_i$  give a measure of the increase in overall image score for a 1-unit increase of the  $i$ -th variable, or



factor, after adjusting for the effects of the other predictors. Therefore, they reveal the importance of each factor for determining the final score of an image.

## Key Results

There were two key scientific results derived from this study, in addition to all the training and development opportunities it generated:

### 1. Ranked list of aesthetic factors

The factor analysis performed as detailed above resulted in a ranked list of aspects affecting human perception of visual aesthetics (in photography) as *determined* by the DPChallenge user community (professional and serious amateur photographers). The final list of factors is shown in the table at the right. Factors appear ordered by absolute value of their regression estimates, which can be interpreted as a natural interval scale to quantify their effect on the perceived aesthetic quality of images. This result has obvious application to image retrieval systems, which can leverage this knowledge to build models for visual features mapping these important aspects from the image dataset.

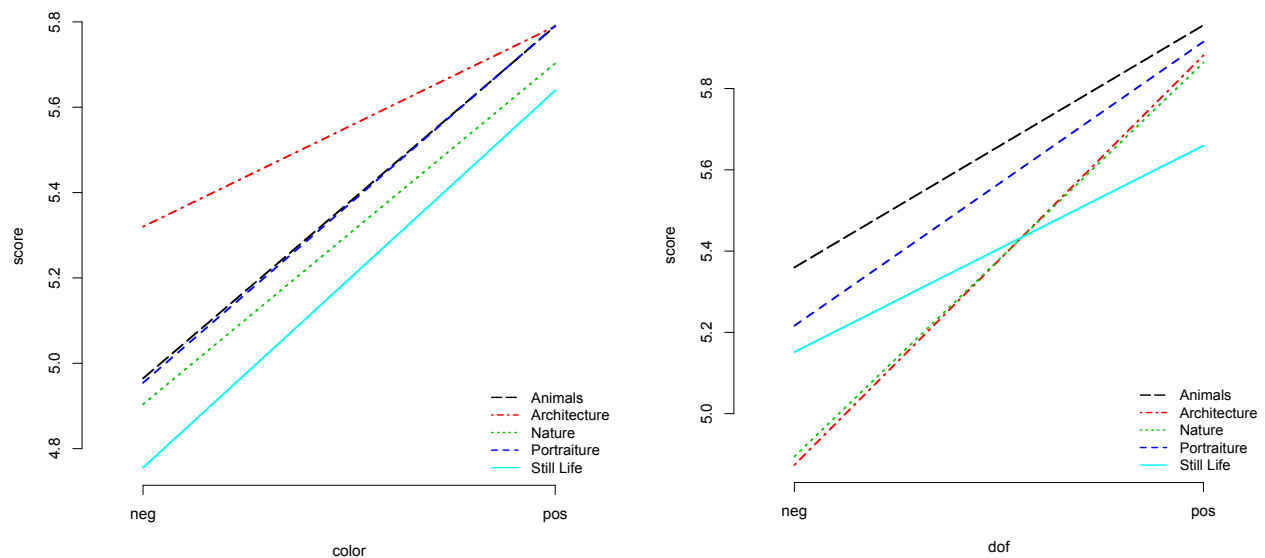
Rank	Factor	Coefficient
1	Execution	0.1422
2	Composition	0.1288
3	Color	0.1049
4	Tone	0.1040
5	Depth Of Field	0.0716
6	Contrast	0.0687
7	Lighting	0.0683
8	Perspective	0.0672
9	Reflection	0.0615
10	Exposure	0.0600
11	Place	0.0545
12	Macro	0.0487
13	Message	0.0473
14	Detail	0.0440
15	Sharpness	0.0405
16	Focus	0.0365
17	Portrait	0.0360
18	Framing	0.0352
19	Pose	0.0335
20	Texture	0.0291
21	Angle	0.0209
22	Idea	-0.0115
23	Eyes	0.0095

### 2. Interaction effect with image category

The fellow also studied how image category had an effect on the list of relevant factors and their order. Noticeable differences were observed between categories, supporting the intuition that factors have a relative effect on the aesthetic value depending on the image subject. The table below shows the top 5 most relevant factors detected for five popular photography categories (nature, portraiture, animals, architecture and still life). Relevant factor differences between these 5 categories are highlighted with boldface font. For instance, in nature photographs, lighting (difficult to control as they are mostly outdoors shoots) and reflections (hinting at the presence of attractive elements, such as water) are highly valued. On the other hand, attributes related to the subject, such as pose, are more determinant to overall appeal for animal photographs. In the case of portraiture, composition and depth of field (which helps isolating the subject from the background) have a higher weight on the overall quality. In the case of architecture, execution (low angles or aerial photographs) is a highly valued attribute.

Nature		Portraiture		Animals		Architecture		Still Life	
Composition	.124	<b>Execution</b>	.149	Composition	.138	Composition	.126	<b>Execution</b>	.132
Color	.111	<b>Composition</b>	.135	Color	.129	Color	.108	Composition	.131
Tone	.107	Color	.089	Tone	.111	Tone	.107	Color	.109
<b>Lighting</b>	.083	Tone	.082	<b>DOF</b>	.096	DOF	.100	Tone	.094
DOF	.082	Lighting	.081	<b>Portrait</b>	.088	<b>Perspective</b>	.100	Lighting	.083

The figure below shows the interaction effect for the *color* and *depth of field* factors across these 5 photography categories. It can be observed that *color* is significantly different for the architecture category, while in the case of *depth of field*, categories can be grouped in three clusters: (animals, portraiture), (nature, architecture), still life.



## Outputs

1. Literature review on NLP methods, sentiment analysis in particular.
2. Study of experimental design and research methods related to factor analysis and aspect rating.
3. Developed software to implement the strategy designed (May to September 2011)
  - a. Java + Python implementation of DPChallenge crawl using Hadoop (fully documented in D2.1)
  - b. Python-based implementation of sentiment analysis algorithm
  - c. Python-based tools to read, parse and pre-process user comments collected from DPChallenge
  - d. Python-based tools to generate and import comment samples for training the supervised sentiment analysis algorithm

- e. Script to perform the factor analysis in R
- 4. Collected dataset including over 600,000 images and associated metadata (dataset characteristics are documented in D2.1)
- 5. Manuscript detailing research written in August-September 2011
- 6. Submitted and accepted into ACM CHI 2012 Conference series in September 2011. Article information: *Your Opinion Counts! Leveraging Social Comments for Analyzing Aesthetic Perception of Photographs* by Jose San Pedro Wandelmer in collaboration with Poonam Suryanarayan (Penn State University).
- 7. Covered one objective of WP3 regarding the use of user generated data to gain high level knowledge of multimedia information, *aesthetics* in this case, circumventing the semantic gap.

## Status

Complete. This work has been submitted and accepted for publication at the ACM International Conference on Human Factors in Computing 2012, CHI'12. In addition, the sentiment analysis tools developed are likely to be considered for transfer into interested business units and teams at the host institution (Telefonica I+D) in the returning phase of the project.

## 1.2 - Analysis of Multimodal Feature Representations of Images for Aesthetic Modeling

In this second study, the fellow considered the use of multimodal feature representations for modeling the visual aesthetic value of images. Both the analytical solution (multimodal statistical models) and the application (image aesthetic modeling) are core to this project proposal.

The aesthetic value of photographs is a very subjective concept, and therefore poses a big challenge in terms of modeling. However, researchers have agreed on a set of principles that are key in the human perception of aesthetics in relation to photographs. High quality pictures tend to exploit shallow depths of field captured using wide apertures, which create photographs with very sharp subjects surrounded by out of focus backgrounds (known as *bokeh*). Composition is also fundamental: specific proportion-related rules (e.g. golden ratio, rule of thirds) are known to produce more appealing images. These rules define the optimal position, size and spatial relations for the main subject and the rest of elements in the photograph. Color (e.g. contrast, vividness) as well as coarseness (e.g. sharpness, texture) features have also direct influence over our perception of visual aesthetics.

Most aesthetic inference methods analyze visual content to determine image quality based on these accepted rules. While they achieve relative success, leveraging contextual information (e.g. user provided tags) outperforms purely visual models. The fellow studied the combined use of both types of information in this application: visual-based features extracted from image content and text-based features extracted from user provided context.

## *Methodology*

The fellow used a rating inference approach to image aesthetic modeling, where context and content features are combined into a hybrid representation to train prediction models of quality scores for photographs. These quality scores serve as proxies for aesthetic value.

In order to model the predictive relationship between aesthetic features extracted from images and aesthetic scores as proposed, it is required to have access to a dataset containing pictures, contextual information and user ratings. The fellow used the DPChallenge dataset described in Section 1.1 for the following main reasons: it is a large-scale representative set, it contains all the necessary elements (images, context and user ratings) and comprises rich and insightful information about the quality of aesthetic value of images in the form of user comments (as shown in Section 1.1).

The fellow considered the following hybrid feature representation of images: text-based features were extracted from user comments using the sentiment analysis tool developed for Section 1.1; relevant visual-based features, reported in previous literature to capture image aesthetics (reviewed in D1.1), were extracted directly from image content. The amount of user feedback available from DPChallenge results in a large annotated dataset of photographs, with multiple users leaving their feedback for the same photo in the form of comments and ratings. Hence, it is expected that the average of these opinions would yield an aesthetic prediction model that reflects the perception of the community.

The fellow obtained hands on experience in the use and tuning of a supervised learning approach to create the aesthetic model. In particular, he focused on learning regression models using SV- $\epsilon$  regression, as indicated in the project proposal. Support Vector Machines (SVMs) are powerful supervised learning methods that allow for building discriminative classification and regression models, both linear and non-linear. SVMs are widely used in the research community and perform excellently well in a wide variety of problems and domains with relatively large datasets.

The fellow could experiment with different feature representation schemes using the features previously described. In particular, he considered a purely visual representation of images (V), a purely textual representation (C), and a combined representation (VC). This process provided insightful results about the possibilities of each of the information sources considered.

## *Key Results*

The analysis of prediction accuracy was performed using three different metrics: R-Squared, Spearman's  $\rho$  and Kendall's  $\tau$ . R-Squared is a widely used metric to test models' goodness of fit based on the aggregated prediction error. Spearman's  $\rho$  and Kendall's  $\tau$  are metrics of rank correlation. They provide a measure of the prediction power of models by looking at rank differences when sorting elements by the observed and the predicted values.

Visual features obtained the predictions with lowest accuracy according to all three metrics. This result is in consonance with previous works in the aesthetic prediction field. The use of contextual information, user comments in this case, achieves consistently higher scores for all accuracy metrics. In contrast to visual-based aesthetic modeling, the comment-based

representation conveys higher level information about the quality of pictures because it is reported by human observers. This result shows that unstructured user comments can be analyzed to create representation that clearly outperforms visual-based models. The proposed hybrid representation led to additional improvements over the comment-based strategy.

	Visual	Context	Hybrid
R-squared	0.0988	0.3726	<b>0.3988</b>
Spearman Rho	0.3133	0.5839	<b>0.6107</b>
Kendall's Tau	0.2125	0.3726	<b>0.4352</b>

### *Outputs*

1. Designed a hybrid feature representation of photographs suitable for aesthetics modeling.
2. Developed Python-based implementation of extractors for the visual features used.
3. Acquired hands-on experience with the use of support vector machines applied to both classification and regression models (using the libsvm tool).
4. Conducted literature review on performance metrics associated to prediction model performance.
5. Manuscript detailing research written in October - November 2011.
6. Submitted and accepted into ACM WWW 2012 Conference series in November 2011. Article information: *Leveraging User Comments for Aesthetic Aware Image Search Reranking* by Jose San Pedro Wandelmer, in collaboration with Tom Yeh (University of Maryland) and Nuria Oliver (Telefonica I+D).
7. Covered two key objectives of WP3 regarding aesthetic modeling: combining visual and context to create a hybrid feature representation, and training models using such representation.
8. Conducted a quantitative evaluation of the proposed representations (visual only, textual only, hybrid) and models, which covers one of the objectives of WP4 in regards to aesthetic modeling evaluation.

### *Status*

Complete. This work has been submitted and accepted for publication into the ACM International Conference on the World Wide Web 2012, WWW'12. This project has generated a number of open research questions, such as how SVMs could compare to alternative learning approaches (e.g. ensembles) as well as the viability of the approach in other application scenarios with noisier user comments.

### 1.3 - Study of Aesthetic Features Applied to the Ranking of Image Search Results

In this third study, the fellow focused on the problem setting of Web image search and investigated the influence of visual aesthetics in the perceived quality of search results. Most of the previous literature on enhancing image search results considers methods aimed at dealing with noisy metadata with the goal of promoting *relevant* content to the top ranks. However, the large size of Web image collections poses an additional challenge: at this scale, the chances of having too many assets similarly relevant to the original query grow. For instance, querying for “dog” would find thousands of relevant images in typical Web image datasets. Increasingly sophisticated ranking and reranking schemes solely based on *relevance* can deal with the problem only to a certain extent. When too many relevant resources exist in the dataset, additional parameters need to be considered for ranking search results.

In this study, the fellow focused on an additional aspect to incorporate to the ranking of image search results: visual aesthetic appeal. The aesthetic appeal of images relates to their ability to generate a positive response in human observers. Such a response can be affected by objective and subjective factors, and is able to create important emotional binds between the observer and the image. The hypothesis that motivated this research is that, when searching for images on the Web, users will tend to prefer aesthetically pleasant images as long as they remain relevant to the original query.

#### *Methodology*

The solution proposed by the fellow takes a list of search results ranked by relevance and rerank them by factoring in aesthetic properties. We call this combination of relevance and aesthetic scores for ranking *aesthetic-aware reranking*. Relevance scores are commonly obtained in image search engines using a text-based retrieval approach to match the query terms with metadata from the images (e.g. tags, title, description). Aesthetic scores can be predicted from the analysis of image content and context. The fellow used the DPChallenge dataset introduced in Section 1.1 as it comprised all of the elements required to perform the study: metadata associated to images to perform the text-based search and generate relevance scores (e.g. title, description), and content and context information suitable for predicting visual aesthetic scores (as described in Section 1.2).

The fellow used Apache Lucene to build an image search system upon the DPChallenge dataset using the standard text-based approach described above, and based on the metadata available from this collection: titles, descriptions, comments and categories. This search engine was then used to generate a list of retrieved images along with their relevance scores for a preselected set of queries. Using this initial result, the fellow predicted two different visual aesthetic scores for each image retrieved: one based on the comment-based prediction model, and a second based on the visual-based prediction model (both introduced in Section 1.2).

In the final stage of the search process, the original text-based relevance scores and the aesthetic values predicted for each image are combined using a linear model following the expression:

$$s(p_i) = \theta_0 r(p_i) + \sum_{j=1}^K \theta_j a^{(j)}(p_i)$$

where  $s(p_i)$  denotes the final combined score for image  $p_i$ ,  $r(p_i)$  its relevance score obtained by the text search engine, and  $a^{(j)}(p_i)$  denotes each of the aesthetic values predicted by the aesthetic regression models.

The resulting images from the search were ranked using three different approaches for evaluating the effectiveness of aesthetic-aware reranking. First, using only text relevance scores, simulating the results that could be obtained from a standard image search engine. Second, combining text relevance scores with visual-based aesthetic scores, that is, using the aesthetic prediction model based on visual features extracted from content. Third, combining text relevance scores with context-based aesthetic scores, that is, using the aesthetic prediction model based on user comments. Both aesthetic prediction approaches have been described in Section 1.2.

An experiment was designed and conducted to assess the performance of each of the ranking strategies described for producing quality image search results for a set of diverse queries. Assessors helped to select their 3 preferred results for a group of 15 images per query using a web-based application. These 15 images had been selected by choosing the top 5 for these three ranking approaches and shuffled to prevent ordering bias. There were 25 different queries to choose from. By analyzing the selection of the assessors we could determine the difference between the different ranking strategies in terms of search results quality.

### *Key Results*

In the aggregated comparison for all queries assessed, the comment-based aesthetic reranking strategy obtained a higher overall performance than the relevance and visual-based strategies. This difference was statistically significant, supporting the hypothesis that users prefer aesthetically pleasant photographs when searching for images. Hence, aesthetic aware rankings, which promote the rank aesthetic images, increase user satisfaction with search results over the original relevance-based ranking. Moreover, the comment-based strategy also performs significantly better than the baseline, aesthetic-aware rank based on visual features. This result is in consonance with the better accuracy performance of comment features discussed in Section 1.2.

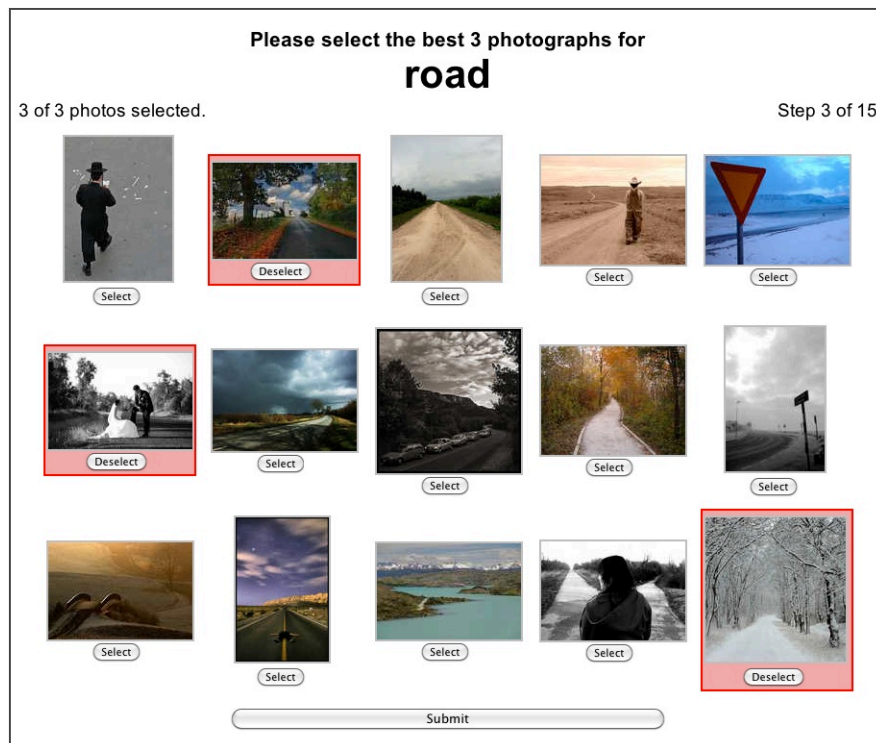
The analysis of performance for individual queries also revealed a clear predominance of the comment-based approach, which obtained the overall highest score in almost 50% of the cases. We also found that users felt more inclined towards aesthetic-aware rankings that combine both relevance and aesthetic scores. In 95% of the cases, an aesthetic-aware reranking was preferred (or not significantly different from the preferred choice). We observed a noticeable inter-query variation of ranking strategy preference, which suggests



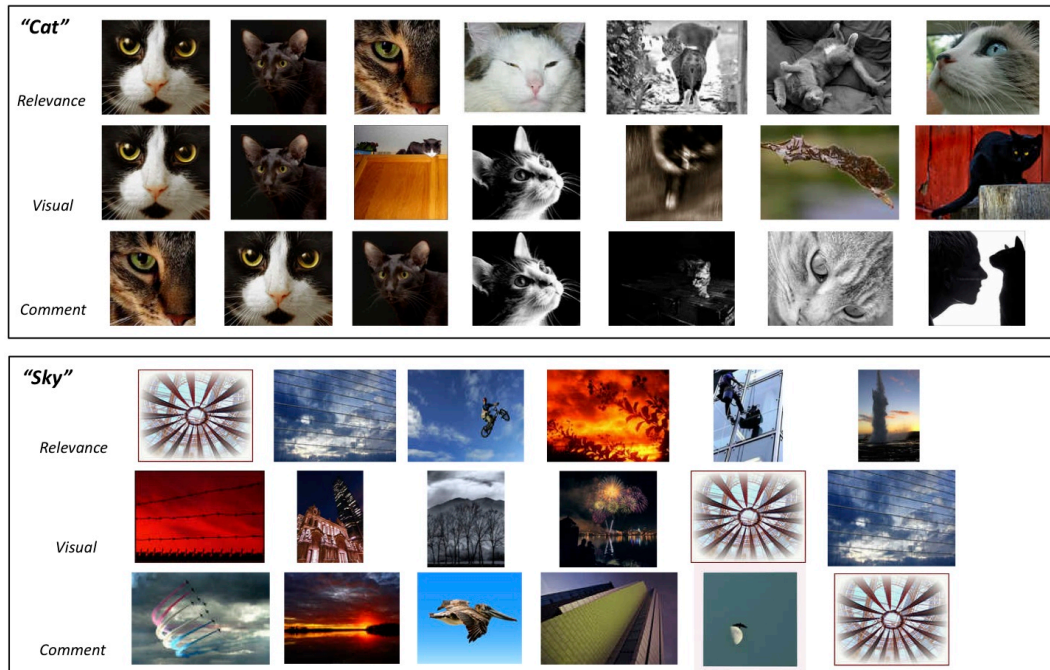
that further optimization strategies should be pursued to adapt the weights of each ranking score to the type of query.

### *Screenshots*

Web interface for the evaluation: Assessors could select their 3 preferred images for the given query using the interface shown below. Users could click on any image to see a full-size version, and could use the buttons below the thumbnails to select/deselect their chosen ones.



Example of images selected using the proposed ranking strategies: The figure below shows the top images retrieved using all three ranking strategies proposed for two different queries (cat, sky).



## Outputs

1. Designed a reranking strategy that combines two orthogonal factors for enhancing image search results: relevance and aesthetics.
2. Acquired hands-on experience on developing an Apache Lucene based search engine to generate relevance scores for images from a given query.
3. Leveraged aesthetic models proposed in Section 1.2 as part of a method to improve information retrieval and user experience with image search.
4. Designed and conducted an experiment to assess the gain in image search results quality.
5. Analyzed the performance of the proposed ranking strategy and showed that it improved image search results quality.
6. Manuscript detailing research written in October - November 2011.
7. Submitted and accepted into ACM WWW 2012 Conference series in November 2011, in combination with the work described in Section 1.2. Article information: *Leveraging User Comments for Aesthetic Aware Image Search Reranking* by Jose San Pedro Wandelmer, in collaboration with Tom Yeh (University of Maryland) and Nuria Oliver (Telefonica I+D).
8. Covered a key objective of the project (across all work packages): demonstrated the ability of the approach proposed in MIESON to combine content and context features to enhance information access in real world applications.
9. Conducted a user evaluation of the ranking approach proposed, which covers one of the objectives of WP4 in regards to aesthetic modeling evaluation.

## *Status*

Complete. This work was submitted along with the study described in Section 1.2 and accepted for publication into the ACM International Conference on the World Wide Web 2012, WWW'12.

## 1.4 – Additional Scientific Achievements

In addition to the three main works presented, the fellow achieved additional scientific results within the timeframe of the project from research he had been carrying out before the project started. While the amount of effort towards these two additional works was marginal during the project, the topics are related to the project and both generated publications after the project start. This section briefly describes their topics and relation to MIESON.

### *A Content-based Propagation Approach Towards Automatic Tagging*

Social multimedia sites are inherently user-centric. Few centralized control policies or systematic ways to organize content exist. They rely on user annotations and feedback, a model that present some drawbacks. Manually annotating content is a time consuming and intellectually demanding process. And not all content receives equal attention from taggers. As a result, annotations are often sparse for large parts of the collection, may lack consistency and present numerous irregularities (e.g., abbreviations and typos).

In the particular case of video oriented social websites an additional problem exists: a vast amount of redundant footage gets uploaded, totalling over 25% of near-duplicate videos detected in search results. Such redundant content can affect retrieval performance, by increasing browsing time to skip repeated entries, or require additional processing to eliminate highly overlapping content.

This work focused on the study of the leading social video sharing platform, YouTube. Rather than viewing redundancy as a problem to be excised from the system, it shows that, by the use of appropriate content-based analysis tools, duplication can be used to understand collective behavior, and point to ways of improving video websites. Specifically, this work considers the use of content-based copy detection tools to establish connections between videos that share duplicate scenes. Such a visual affinity networks can be used to propagate annotations between neighbor elements, allowing spreading community knowledge in a novel approach to *automatic tagging*. Uploaders of overlapping sequences of the same video provide their personal perspective on its content in the form of different annotations.

This work relates to MIESON in two key aspects. First, it provides an extensive empirical analysis of duplication and dependencies between video overlaps and metadata in YouTube. This analysis reveals insights about the community and how they share video on the web, which is important for the design of efficient information access systems. Second, it proposes an automatic tagging propagation scheme based on the methodology proposed in MIESON: fusion of content (video duplication detection) and context (tags) features.

## *A Case for Multimodal Queries in Information Search*

The multimedia information retrieval community has dedicated extensive research effort to the problem of content-based image retrieval (CBIR). However, these systems find their main limitation in the difficulty of creating pictorial queries. As a result, few systems offer the option of querying by visual examples, and rely on automatic concept detection and tagging techniques to provide support for searching visual content using textual queries.

In this work, a study of a practical multimodal web search scenario is conducted. A system for searching computing articles, which tend to be richly illustrated by screenshots, is analyzed. In it, rich multimodal information queries can be generated to retrieve information, using an intuitive and user friendly approach to CBIR. The system leverages visual correspondences between screenshots as well as text similarity to improve ranking of retrieved articles. Users can naturally create pictorial queries simply by taking a screenshot of the application to retrieve a list of articles containing a matching screenshot.

This work relates to MIESON in the key aspect of sharing the same methodological approach to improving access to multimedia information: combining content and context. The strategy proposed considers a scenario that complements the objectives of MIESON. While MIESON considers the use of multimodal information for representing and understanding content, this work uses this idea to endow users with more powerful tools to search collections.

## 2 - Summary of Training and Transfer of Knowledge Activities

### 2.1 Courses to Acquire New Skills Relevant to the Project

During the outgoing phase of the project the fellow carried out a large number of training activities. The following list enumerates courses taken to learn the set of skills defined within the scope of project MIESON.

#### 2.1.1 [Cloud Computing] Cloudera's Hadoop Developer Training

*Washington DC, USA - August 22<sup>nd</sup> to 24<sup>th</sup> 2011 – 24 hours*

Cloud computing has been identified as an important technology to be learnt within the scope of the project in order to deal with user data generated in online websites, which tend to be large-scale. The cloud computing paradigm has been adopted by the main players in the Internet industry, including companies such as Google, Facebook or Yahoo. This wide adoption, along with the paradigm desirable properties for building scalable systems, supports its relevance within the scope of the project and its value as a set of skills to be acquired by the fellow.

The fellow enrolled in one of the most prestigious training courses in the cloud computing area. Cloudera's *Hadoop Developer Training* provides attendants with the necessary

knowledge of the current cloud computing ecosystem to start working with the technology as soon as the course is finished. The core of the course is Hadoop, an open source implementation of the MapReduce technology invented by Google to create a scalable solution to the Web indexing problem. Cloudera has extensive experience on the use of cloud computing, and is one of the major contributors to the Hadoop open source project.

After the course, the fellow took and successfully passed the certification exam offered by the company: “Cloudera Certified Developer for Apache Hadoop”. The fellow has been able to obtain hands-on experience on the use of the technology learnt during this course to generate one of the datasets used within the project, as described in D2.1. It is expected that these techniques will be used during the next stages of the project.

### 2.1.2 [Statistical Data Analysis] Statistical Research Analysis with R

*Online - October 7<sup>th</sup> to December 8<sup>th</sup> 2011 – 28 hours*

R is an open source statistical programming language widely used in experimental sciences to perform statistical analysis of data. R has experienced an important growth in use in the last years thanks to its adoption by an increasing number of scientists interested in analyzing big sets of data. The community of active supporters is extensive, both contributing new code and libraries, as well as documentation, tutorials and answers to frequently asked questions.

The fellow enrolled in an online course offered by the Georgia R School. Classes were taught in a series of live, interactive, synchronous, online sessions of 3 hours using the eLearning platform Blackboard Collaborate once per week for nine weeks. The instructor was Dr. Geoffrey Hubona, an experienced quantitative researcher, and a faculty member of the Information Systems department at Virginia Commonwealth University.

This course included two different parts. The first part was dedicated to basic elements of statistical research analysis of data, touching on topics such as analysis using data plots, simple inference, conditional inference, analysis of variance, linear regression, logistic regression and generalized linear models. The second part considered more advanced analysis topics, including survival analysis, smoothers and generalized additive models, analysis of longitudinal data, linear mixed models and principal component analysis. The fellow has used R in several of his research works so far, and plans to continue using it in the remaining stages of the project.

### 2.1.3 [Machine Learning] Stanford’s Machine Learning Online Course

*Online - October 9<sup>th</sup> to December 6<sup>th</sup> 2011 – 18 hours + lab exercises*

Machine learning tools at the core of most technologies related to information management and access applications, from visual information analysis to multimedia content retrieval. These methods allow the computer to learn from existing data and build models useful for a variety of tasks. These models are extensively used within the scope of the MIESON project.

The fellow enrolled in this online course taught by the Director of the Artificial Intelligence Lab at Stanford University’s, Prof. Andrew Ng. Ng is the author of over 100 published papers

in machine learning, and his work in learning, robotics and computer vision has been featured in a series of press releases and reviews.

This course has enabled the fellow to learn the fundamentals of the most commonly used machine learning methods as well as the best practices in real world use of machine learning methods. While the fellow was familiar with the machine learning field, this course has widely enrich his understanding of the tools and will allow him to use them more efficiently in the remaining stages of the project. The fellow was awarded a statement of accomplishment for the timely completion of lab exercises and review tests assigned during the course.

#### 2.1.4 [Statistical Data Analysis] Programming, data analysis and statistics with R

*University of Maryland, College Park, MD, USA – January 11<sup>th</sup> to January 20<sup>th</sup> 2012 – 20 hours*

This R course focused on fundamental concepts of the R programming language, including importing and exporting data, working with data frames, aggregating and summarizing data, basic plots, data analysis and statistical data modeling. The course also included an advanced track focused on data visualization for researchers using R, and in particular using the ggplot module, which widely extends R graphic capabilities, to create plots that tell a concise story for other people and to polish up plots for presentation and publication. The course was part of the Winter Storm Workshop organized by the University of Maryland's language science community, a free to access event open to all researchers interested in statistical data analysis. The advanced track was taught by Scott Jackson, Assistant Research Scientist at The Center for Advance Study of Language, while the basic R course was taught by PhD students Ewan Dunbar, Shevaun Lewis and Annie Gagliardi.

This workshop also included presentations focused on career development in academia, including:

- Securing a job in academia, by Prof. Susanne Jaeggi: a full review of how academia hiring works in the most prestigious academic centers in the world, with emphasis on how to apply for academic jobs (preparing cover letter, research statement and CV), how to be successful at the interview process, and how to negotiate conditions once you are offered the job.
- Getting ready for the job market, by Prof. Howard Lasnik: focused on helping scientists get prepared for the job market, either in academia or industry, including advice for preparing the most important materials required (CV, etc), resources to find jobs (listings, contacts, etc), and a extensive list of tips to be successful in the task of finding a job in the scientific research field.



## 2.2 Tutorials

In the tutorials section, we describe a number of short courses (8 hours or less) taken by the fellow and focused on very specific topics relevant to the project and the fellow's career development plan.

- [Research Methods] Practical Statistics for User Research

*Vancouver, British Columbia, Canada - May 10<sup>th</sup> to 11<sup>th</sup> 2011 – 5 hours*

*Taught by: Jeff Sauro (Oracle, Measuring Usability LLC), Jim Lewis (IBM)*

Tutorial on experimental design, and qualitative and quantitative methods for conducting user-oriented research. Part of the ACM International Conference of Human Factors in Computing 2011 (*CHI'11*).

This tutorial on usability analysis and user-research focused on learning how to conduct and interpret appropriate statistical tests on small and large sample usability data. It covered topics such as A/B Testing, determining sample size for comparing designs and benchmarking studies, understanding the limits from sample usability data through use of confidence intervals, and practicing what statistical test to perform and how to interpret the results.

- [Multimedia analysis] Frontiers of Multimedia Research

*Scottsdale, Arizona, USA - November 28<sup>th</sup> 2011 – 4 hours*

*Taught by: Alan Hanjalic and Martha Larson (Multimedia Information Retrieval Lab, Delft University of Technology)*

Tutorial on the present and future challenges of multimedia information retrieval. Part of the ACM International Conference on Multimedia 2011 (*MM'11*).

The aim of this tutorial was to provide insights into the most recent developments in the field of multimedia retrieval and to identify the issues and bottlenecks that could determine the directions of research focus for the coming years. The tutorial presented new lines of work that consider multidisciplinary approaches including information retrieval, speech and language processing and network analysis. Also, it discussed the evaluation of new algorithms making use of recent tools such as crowdsourcing.

- [Information Retrieval] Design and Analysis of Large Scale Log Studies

*Vancouver, British Columbia, Canada - May 9<sup>th</sup> 2011 – 4 hours*

*Taught by: Susan Dumais (Microsoft Research), Robin Jeffries (Google), Dan Russell (Google), Diane Tang (Google), Jaime Teevan (Microsoft Research)*

Logging user behavior is a crucial aspect of understanding how people actually use online systems. This tutorial gave a practical overview of the issues involved in obtaining,



cleaning, interpreting and analyzing large log data sets, with application to evaluate the performance of ranking methods. Part of the ACM International Conference of Human Factors in Computing 2011 (*CHI11*).

- [Social Computing] Social Media Networks & Communities

*College Park, Maryland, USA- May 24<sup>th</sup> 2011 – 2 hours*

*Taught by: Jen Golbeck (University of Maryland), Derek Hansen (University of Maryland)*

Tutorial on the topics of network topology analysis and characterization, analysis of social metadata and social popularity metrics. Part of the HCIL Symposium 2011 at the University of Maryland.

Social media have created a wealth of social data in the form of Facebook friends, Twitter Followers, hyperlinks, and shared tags. This tutorial introduced network analysis concepts and tools for the analysis of social media and online community interactions. Prof. Golbeck and Hansen discussed techniques and tools that help to harvest and make sense of this relational data. The tutorial included a mixture of instruction and discussion. The free tool NodeXL (developed by Prof. Hansen) was introduced and demoed, along with methods for understanding and visualizing relationship types and strengths.

- [Management] Entrepreneurship

*College Park, Maryland, USA- May 24<sup>th</sup> 2011 – 2 hours*

*Taught by: Ed Barrientos (CEO Brazen Careerist)*

Tutorial on the topic of transfer of knowledge and how to turn scientific research into business. Part of the HCIL Symposium 2011 at the University of Maryland.

The tutorial introduced several aspects involved in the creation of technological-based start-ups and focused on financial matters of the transfer of technology endeavors, such as raising capital to fund a start-up, a very challenging task in today's economic environment. Barrientos talked about the difference between angel investors and venture capitalists, under what circumstances should entrepreneurs consider one or another, and finally *must-have* ingredients for a compelling investor pitch.

## 2.3 Attendance to Seminars

The fellow has attended a number of talks given by prominent figures of the information science and technology fields. These talks have provided him with the opportunity of getting to know about interesting research ideas and challenges relevant to project MIESON that other researchers and technologists are currently working on:

1. *Emerging trends in Search User Interfaces* (Prof. Marti Hearst, Professor in the School of Information at The University of California, Berkeley): The talk focused on analyzing the potential future direction of search user interfaces, discussing some important trends in the use of information technology and how these may affect search in future. This includes a notable trend towards more "natural" user interfaces, a trend towards social rather than solo usage of information technology, and a trend in technology advancing the integration of massive quantities of user behavior and large-scale knowledge bases.
2. *Out of the Ivory Tower: Lessons learned at Twitter* (Dr. Jimmy Lin, Assistant Professor at The University of Maryland, College Park): Dr. Jimmy Lin talked about his experience as a visiting research fellow in Twitter. Dr. Lin had the opportunity to work in a team of engineers developing information retrieval tools that are running in production on Twitter's main site. His talk provided a great overview on how experimental research developments are carefully engineered to be deployed in real world applications in the context of a fast-paced startup with short time-to-market product cycles.
3. *Large-Scale Social Media Analysis at Facebook* (Dr. Rong Yan, Research Scientist at Facebook): In this talk Dr. Rong Yan overviewed a number of Facebook algorithms and applications used for different purposes such as: identifying popular trends, discovering latent topics of users, and detecting user happiness. Related Facebook applications in the domain of media such as face detection and video storage were also discussed.
4. *Free Software in Ethics and in Practice* (Richard Stallman, Founder of the Free Software Foundation): Richard Stallman talked about the differences between free and privative software, both from an ethical and practical standpoint. In particular, Stallman discussed concerns on the use of applications in the cloud, which may introduce important shortcomings in terms of privacy and user rights. Given the increasing popularity of such online applications, they are likely to require regulations that help protect users without limiting the positive effect of technology for society.
5. *Making Topic Models More Human(e)* (Jordan Boyd-Graber, Assistant Professor, College of Information Studies, University of Maryland, College Park): Topic models, which automatically discover the themes which permeate a corpus, are a popular tool for discovering what's being discussed. However, topic models aren't perfect; errors hamper adoption of the model, performance in downstream computational tasks, and human understanding of the data. However, humans can easily diagnose and fix these errors. Jordan Boyd-Graber presented a statistically sound model to incorporate hints and suggestions from humans to iteratively refine topic models to better model large datasets. He also examined how topic models can be used to understand topic control in debates and discussions. We demonstrate a technique that can identify when speakers are "controlling" the topic of a conversation, which can identify events such as when participants in a debate don't answer a question, when pundits steer a conversation toward talking points, or when a moderator exerts her influence on a discourse.
6. *Unsupervised Textual Analysis with Rich Features* (Vlad Eidelman, PhD Student, Computer Science, The University of Maryland): Learning how to properly partition a set of documents into categories in an unsupervised manner is quite challenging,

since documents are inherently multidimensional, and a given set of documents can be correctly partitioned along a number of dimensions, depending on the criterion. In this talk, Vlad presented a feature-enhanced unsupervised model for categorizing textual data. The presented model allows for the integration of arbitrary features of the observations within a document. While in generative models the observed context is usually a single unigram, or bigram, the presented model can robustly expand the context to extract features from a block of text of larger size.

## 2.4 Conferences/Events Attended

1. ACM CHI 2011 (Vancouver, BC, Canada) May 7<sup>th</sup> to 12<sup>th</sup> 2011
2. HCIL Symposium 2011 (College Park, MD, USA) May 25<sup>th</sup> to 26<sup>th</sup> 2011
3. ACM Multimedia 2011, (Scottsdale, AZ, USA) November 28<sup>th</sup> to December 1<sup>st</sup> 2011
4. Winter Storm Workshop (College Park, MD, USA) January 11<sup>th</sup> to 20<sup>th</sup> 2011

## 2.5 Publications

The fellow published one full conference paper and one journal paper during the outgoing phase.

Tom Yeh, Brandyn White, **Jose San Pedro**, Boriz Katz, and Larry S. Davis. (2011). *A case for query by image and text content: searching computer help using screenshots and keywords*. In Proceedings of the 20th international conference on the World Wide Web, WWW '11. ACM, New York, NY, USA, 775-784

**Jose San Pedro**, Stefan Siersdorfer, and Mark Sanderson. 2011. *Content redundancy in YouTube and its application to video tagging*. ACM Transactions on Information Systems. 29, 3, Article 13, 2011.

Two additional papers about the research carried out during the outgoing phase have been accepted for publication in spring 2012 into the following two top-tier conferences: ACM International World Wide Web Conference (WWW'12) and ACM International Conference on Human Factors in Computing (CHI'12). They are described in the “Dissemination Activities” section of this report.

## 2.6 Participation in Research Events

The fellow was invited to be a panelist for the College of Information Science and Technology Symposium of The Pennsylvania State University on March 25<sup>th</sup> 2011. The panel, entitled “Doing Internet Research. Methods, Challenges and Large Scale Data Sets”, considered the

discussion of ideas related to the new methodological approaches needed for researching data available in The Internet, from the subtleties of online human interactions to the huge amount of raw data we can now access.

The fellow was also part of the Program Committee for IEEE International Conference on Machine Learning and Applications 2011.

## 2.7 Mentoring

The fellow has had the opportunity to collaborate with PhD students of the outgoing host institution, giving them feedback and ideas about their research in the regular group meetings.

## 2.8 Internal/External Talks

The sentiment analysis technology developed (Section 1.1) is relevant for mining the opinion of users in social networks, an increasingly popular feedback channel where users complain about or recommend products, services and companies. This vast set of community feedback enables stakeholders to understand users' feeling about their and their competitors' products, providing opportunities to react to users demands in a timely manner. The fellow gave a talk explaining in more detail the possibilities of the technology developed within the scope of the objectives of the host institution, Telefonica I+D, in January 2011.

## 3 Summary of Clearly Significant Results

- The fellow achieved the objectives defined in the work program for the outgoing phase.
- The fellow published two papers:
  - Tom Yeh, Brandyn White, **Jose San Pedro**, Boriz Katz, and Larry S. Davis. (2011). *A case for query by image and text content: searching computer help using screenshots and keywords*. In Proceedings of the 20th international conference on the World Wide Web, WWW '11. ACM, New York, NY, USA, 775-784
  - **Jose San Pedro**, Stefan Siersdorfer, and Mark Sanderson. 2011. *Content redundancy in YouTube and its application to video tagging*. ACM Transactions on Information Systems. 29, 3, Article 13, 2011.
- Two additional papers have been submitted and accepted for publication in spring 2012, at the beginning of the returning phase.

- The fellow attended an important number of talks, tutorials and courses that provided him with research skills and techniques relevant to the project and defined in his career development plan.
- The fellow gave a talk at the host institution offices to explain additional details and possibilities of the sentiment analysis technology developed (Section 1.1) for the host institution.
- In addition to publications, the fellow: interacted with several research groups at the outgoing host institution, mentored students, gave several talks and was an invited panelist at the IST Symposium. All these activities helped him develop his communication skills as defined in his career development plan.
- The fellow attended several scientific research events and conferences that provided him with a number of networking opportunities in consonance with his career development plan.

## 4 Deviations

The following section details the deviations from the original work plan defined in the proposal.

The most significant deviation from the original project proposal has been the reduction of the outgoing phase to 14 months from the 24 initially scheduled. This reduction has been already discussed with the project officer. The Grant Agreement (Annex I) has been modified in the 2<sup>nd</sup> amendment in order to reflect these changes in the original workplan that has been already approved by both the host institution and the project officer. This report reflects the objectives as they are described in the most updated version of the Grant Agreement.

This deviation originates from the difficulties arisen during the negotiation with the outgoing host institution on the conditions ruling the short stay of the fellow at The University of Maryland, as stated in the 1<sup>st</sup> amendment of the Grant Agreement approved in November 2011. The outgoing host institution did not facilitate the paperwork necessary to carry out such short stay. Difficulties in the communication between the fellow and the outgoing phase host made continuing with the outgoing phase not feasible.

As reported in this document, the fellow has achieved most of the objectives originally planned for the outgoing phase, and has completed the most important training activities arranged, despite the shortening of the outgoing phase.

One additional minor comment regarding the original work plan. Licensing and copyrights prevents us from releasing the dataset collected during this first year. This mainly affects D2.2 expected at the end of month 24. This circumstance was already foreseen in the project proposal, so it is not really a deviation from the plan.

## 5 Explanation of the Use of the Resources

The fellow has been working on MIESON project full-time during the period February 2011 – March 2012. The amounts received by the fellow have been entered on the NEF in Form C section. The following table gathers other project **actual costs** for the above mentioned period related to participation expenses.

Other major cost items for MIESON project for the period February 2011 – March 2012			
Work Package	Item description	Amount in € (2 decimals)	Explanations
1, 2, 3, 4	Participation Expenses	6.825,84 €	<p>Travel to attend ACM CHI 2011 conference (06/05/2011) (Vancouver/ Canada) = 1075,6 €</p> <p>Travel to attend ACM Multimedia conference (28/11/2011) (Scottsdale/ Arizona) = 297,88 €</p> <p>Internal MIESON Project at TID facilities (08/01/2012) (Barcelona/Spain) = 252,49 €</p> <p>2011 Human-Computer Interaction Lab Symposium Fee = 183,05 €</p> <p>Cloudera Developer Training for Apache Hadoop Fee = 1.680,40 €</p> <p>Using R For Statistical Research Analyses Training = 362,44 €</p> <p>ACM Multimedia 2011 Fee = 545,49 €</p> <p>ACM CHI 2011 Fee = 904,27 €</p> <p>ACM Conference On The World Wide Web Fee = 740 €</p> <p>ACM CHI 2011 Conference Fee = 595,22 €</p> <p>"Modern Information Retrieval" book = 189 €</p>
TOTAL COSTS		6.825,84 €	

## ADDITIONAL INFORMATION

There is no additional information to be provided.



## DISSEMINATION ACTIVITIES

The following describes all dissemination activities carried out during the outgoing phase as well as the dissemination activities planned for the returning phase at the time of writing this report.

Two scientific articles have been published so far, a journal article in ACM Transactions on Information Systems (high impact factor) and a full conference paper at ACM Conference on the World Wide Web 2011:

- Tom Yeh, Brandyn White, **Jose San Pedro**, Boriz Katz, and Larry S. Davis. (2011). *A case for query by image and text content: searching computer help using screenshots and keywords*. In Proceedings of the 20th international conference on the World Wide Web, WWW '11. ACM, New York, NY, USA, 775-784
- **Jose San Pedro**, Stefan Siersdorfer, and Mark Sanderson. 2011. *Content redundancy in YouTube and its application to video tagging*. ACM Transactions on Information Systems. 29, 3, Article 13, 2011.

Two additional articles have been submitted to conferences and have been accepted for publication during spring 2012 (ACM Conference on the World Wide Web 2012 and ACM Conference on Human Factors in Computing, CHI, 2012). All these conferences are top-tier events and extremely competitive in terms of acceptance rate.

- **Jose San Pedro**, Poonam Suryanarayan. (2012). *Your Opinion Counts! Leveraging Social Comments for Analyzing Aesthetic Perception of Photographs..* In Proceedings of the ACM International conference on Human Factors in CHI '12, Austin, TX, USA.
- **Jose San Pedro**, Tom Yeh, and Nuria Oliver (2012). *Leveraging User Comments for Aesthetic Aware Image Search Reranking*. In Proceedings of the ACM International World Wide Web Conference, WWW'12, Lyon, France.

During the last two months of the outgoing phase, the fellow has dedicated a considerable amount of time to preparing for the presentation of these two works in the aforementioned forums, including: write-up of the camera-ready version with all changes suggested by reviewers to the original submission, composition of the slides for both talks (WWW and CHI), registration and trip logistics.

In addition to these dissemination activities, the fellow is planning to attend the biannual Marie Curie event collocated with the European Science Open Forum (ESOF), to be held this year 2012 in Dublin. This event provides a great opportunity to talk about the goals of Marie Curie projects with other fellows at a similar point of their careers.

## PROJECT MANAGEMENT

To date, the fellow has completed the work planned for the outgoing phase as reflected in the last amendment to Annex I of the Grant. Three different studies regarding multimodal feature representations, multimedia understanding and information access regarding the topic of aesthetic modeling have been carried out, leading to two publications accepted in top-tier conferences in 2012. Results obtained so far support the validity of the approach proposed in project MIESON. The most important scientific objective of the outgoing phase, learning the methodological approach towards using content and context feature representation of multimedia information, has been accomplished

The fellow has advanced on the scientific objectives of the proposal as well as in the training tasks defined in his career development plan. He has published 2 scientific papers at top-tier peer-reviewed international conferences; attended numerous conferences, courses, tutorials and talks; participated in networking events; and enhanced his communication skills by giving one talk at each host institution and participating as an invited panelist in a symposium. He has also finished the reporting tasks associated to the fellowship in a timely manner, which include the deliverables for WP1 and WP2.

Regarding deliverables, the fellow completed both D1.1 (Data Representation Schemes) and D2.2 (Datasets) as planned, covering aspects about ways to collect and represent data from social media websites required for subsequent work packages, specifically WP3 and WP4.

During the returning phase, the fellow will continue pursuing the scientific objectives of the proposal. In addition, he will engage in at least two additional networking and communication activities related to the presentation of his 2 accepted papers in top-tier international conferences in 2012. He will continue his training in skills highly relevant for the project, including cloud computing, natural language processing, and machine learning. Given the current status of the project, we do not anticipate any problems in completing the work program proposed.

The main deviation from the original plan has been the reduction of the outgoing phase, from 24 to 14 months. This deviation originates from the first amendment introduced to the Grant Agreement on November 29<sup>th</sup>, which added a stay away from the outgoing host institution for a period of 6 months at the University of Maryland's Institute for Advanced Computer Studies (UMIACS). Such stay would have covered training in theoretical and practical aspects of Cloud Computing in consonance with the original spirit and objectives defined in the project proposal. The fellow encountered unexpected bureaucratic problems with the original outgoing phase institution, which delayed and finally prevented the fellow from completing the short stay. Given that the communication with the outgoing host had not been easy, both the fellow and the host institution decided that it would be best for the fellow and the success of the grant to shorten the duration of the outgoing phase to 14 months. This decision was supported by the fact that the main scientific objectives that had been planned for the outgoing phase had been achieved. This issue was immediately communicated to the project officer, who understood the situation and was supportive of the proposal to shorten

the outgoing phase. Annex I has been modified (2<sup>nd</sup> amendment) to reflect this reduction in the outgoing phase duration and objectives.

In terms of gender and ethical issues, there are none to highlight.