# Expert assisted exploration of photographs

Supporting Users in Exploring Visual Media through Subjective Aesthetic Attributes and Crowd-sourced Tags

A dissertation submitted to the University of Dublin, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

Meltem Gürel

2009

# Declaration

I, the undersigned, declare that this work has not previously been submitted to this or any other University, and that unless otherwise stated, it is entirely my own work.

---

Meltem Gürel

Dated: September 11, 2009

# Permission to Lend and/or Copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

Meltem Gürel

Dated: September 11, 2009

# Acknowledgements

I want to express my genuine gratitude to my supervisor, Dr. Owen Conlan, for all the time spent assisting me and for all the encouragement and guidance offered. I would also like to thank Cormac Hampson for all his work and useful feedback, and Dr. Vincent Wade for the valuable suggestions.

To my family and friends, and especially my fellow NDS classmates, I would like to thank you for being there even at the last minute.

**Meltem Gürel**

*University of Dublin, Trinity College*

*September 2009*

# Abstract

While digital technology extinguished the obligatory use of the photographic film as well as the time consuming chemical photograph development techniques, the number of taken photographs has rapidly increased. As the computer harddisks as well as the photography websites substituted the photo albums, finding photographs from digital archives became difficult. Conventional image retrieval systems that seek to bridge the semantic gap, optimize photograph discovery based on associated tags or their low-level features, which can define the information regarding the content of a photograph, however not the photographs' aesthetic value.

This dissertation investigates the possible benefits of augmenting traditional query techniques with subjective expert knowledge based on the manipulation of raw low-level data in order to empower users in exploring visual media. A novel tool, $X^2$*Photo* is presented which aims to enable users in retrieving photographs from large collections using not only objective tags but also subjective expertise based on photographs' color space.

Evaluation results showed that conventional tag-based systems ignore the appreciative expressions hence limit the users to search via tagged simplifications of a photograph rather than the aesthetics of the photograph itself. Injecting expert knowledge into a conventional system that only offers tag-based searching, allows users to freely express both the aesthetics of the photograph they want as well as the picture it conveys.

$X^2$*Photo* provided users an alternative pathway to access their large photograph collections. Initial user tests showed promise and indicate that this approach can be used to grant users the freedom they seek in relation to photograph discovery.

# Contents

# List of Figures

# Listings

# Chapter 1

# Introduction

"Shooting" a "quality" photograph has been tremendously abridged, especially during processing due to the digital revolution which has had a great effect on the photographic medium. Although one can argue about its advantages and disadvantages, digital cameras (or devices integrated with their capabilities such as cell phones) and the fast decline in their prices, has made photography more popular than ever among amateurs and professionals alike. Owing to the fast advances in digital technology we can now take over 60 pictures per second. But with this rapid growth in personal photograph collections comes the difficulty of finding the right photo. How do we search within millions of "shots"?

## 1.1 Motivation

Many popular image search engines are based on keywords or terms associated with or assigned to a photograph: "tags", which may be inaccurate and misleading at times [Cui et al., 2008]. While they are useful in defining the "content" of a photograph, their single dimensional nature does not allow for precise refinement of photographs. As tags are objective pointers they do not embody aesthetic values, hence are weak for aesthetic appreciation of photographs. Web image search engines that depend on classic tagging systems have no clear information about the meaning or semantics of each individual tag, causing unexpected or irrelevant results to be returned to the user. As tags are coherent among the society and during the period that they were created, their use may not be easily applicable to photography. When a photograph is tagged as “vigorous” one might ask “vigorous to whom?” or “how vigorous?” or even the question “Define vigorous” may arise. As access to collective knowledge is becoming more and more important should search pointers be solely based on such tags which may differ from one person to another?

Content based image retrieval (CBIR) systems can find images that are similar in content to a query image, but providing the initial image may be a problem for the user - browsing through albums can be very time consuming hence inefficient when the collection has a vast number of photographs - or the query might produce a limited result set as the user may not be interested just in the content of the image but its value as a photograph. [Eakins et al., 2004] also points out the posibility of overspecifying the search: e.g. an image containing a car with passengers inside it is submitted, then even if the user is only interested in the car the passengers become part of the query. In the case where a user tries to find photographs with a similar “feel” to the query image, CBIR systems which are generally based on features like texture, color, and shape will return photographs of similar low-level characteristics. Searches based exclusively on this approach may prevent the user from finding the “perfect” photograph since many photographs which could have had the same impact won’t be in the set due to not having similar low-level characteristics. Users may be interested in the overall effect/value of a photograph which can be thought of as the certain mix of raw low-level features.

In accordance to these issues with search engines that rely on traditional data querying techniques, end-users are prevented from utilizing the full potential of large photograph collections. Introducing the use of expert knowledge into the exploration of photographs can enable end-users to discover more accurate results and can also guide them in the whole process by bringing in domain specific identifiers for individual photographs. The knowledge of domain experts can lead the end-users to find photographs through the experts’ vocabulary which may not have been the obvious approach to the end-user initially in defining the photograph they were seeking. Domain experts have clear, heavily subjective perspectives when it comes to defining and evaluating a source of query.

There is an immense amount of literature dedicated to the subject of wine. When analysing a particular type of wine, different perceptions such as color/clarity, odor and taste are used by experts of this domain to express their feelings through descriptive words or phrases. The wine-tasting terminology comprises of high-level subjective terms which are derived from low-level characteristics of a wine. For example, the term "bitter" actually refers to the tannin content of a wine, and the term "oily" is used for the combination of high glycerin and slightly low acid content [Jackson, 2002]. This expert terminology creates a semantic space for wine tasting. The words used can differ among experts but are based on the same characteristics. Applying such subjectivity to the photography domain through multiple disciplines that have an effect on the overall feel of a photograph can help refine a more clear-cut search environment. Subjective expertise can aid and leverage image search in order to allow end-users to explore collections with an awareness.

[Surowiecki, 2004] strongly proposes that a crowd is wiser than, or is as wise as, the smartest person in that crowd, and should definitely be smarter than an expert because "information isn't in the hands of one person. It is dispersed across many people." According to this idea one can say that crowd sourcing really works well when the sum of the crowd possesses more knowledge than the domain experts. However, when applied to the photography domain, apart from defining the "content" of the photograph can average enthusiasts have a well-formed idea about the photograph itself? Would they be able to express their query in precise wording let alone define the actual photograph itself? Since the judgment and likes of the crowd will be widely varying, their perception of a particular photograph may be, on the contrary, misleading and to tap into that "wisdom" may be a mistake. One should also not forget the power of influence in a crowd: if the photograph were to be defined and described repetitively within the same context with similar annotations it might lead to average and common views which would be recognized by most, neither offending nor exquisite.

Based on the assumption that many people react with pure emotions rather than knowledge when it comes to describing a photograph, the outcome may not be a rational decision. Here a domain expert might be more insightful in pointing out the characteristics of an individual photograph, bearing in mind not how a photograph is but the effect it may have on one. Also taking into account that the raw low-level meta-data of a photograph is already accessible by the expert, is there a strong necessity for widespread knowledge? Can the highly subjective perspective of the expert suffice in finding the perfect photograph?

## 1.2 Research Question

*To what extent can the combination of tags and subjective expertise support end users in exploring visual media?*

Here we refer to crowd sourced annotations as tags and particularly those assigned by Flickr users to their digital photographs. Subjective expert knowledge is based on the manipulation of non-textual low-level data contained within a digital photograph; specifically the hue, saturation and lightness of its dominant colors. We propose that when combined, these two features should enable exploration through both content, achieved via the selection of tags, and through an aesthetic perspective derived from the expert knowledge. This research aims to investigate the accuracy and efficiency of the result set produced via the proposed approach.

The aims of the research can be listed as follows:

- Determining authentic semantic information to visual characteristics of an object, e.g. determining the "warmth" of color in a photograph
- Facilitate exploration using aesthetics
- Utilizing semantic information as a means to prioritize and organize visual objects in order to render them in an intuitive and accessible manner.
- Provide end-users with an alternative access pathway when browsing for photographs.

## 1.3 Objectives

In order to answer the research question brought up above, the following objectives were identified:

- Research the current state of the art in image retrieval systems in order to derive the requirements for a system that enables end-users in exploring visual media.
- Following an iterative method, create an expert vocabulary based on the processing of low-level data to produce semantically relevant image attributes and identify a framework combining this domain expertise with crowd sourcing.
- Develop the final tool which will empower users to explore large volumes of visual data using semantic criteria to refine and focus their needs.
- Perform user-centric usability and acceptance tests to evaluate the effectiveness of the tool as a means to visual search and exploration.

## 1.4 Approach

To complete the objectives and answer the research question posed, we first researched image retrieval systems, possible photography websites to gather data from and their APIs. Deciding on Flickr, we cached over 12000 public photographs along with related Exif metadata. Next, we designed and developed a prototype application in PHP to start building an expert vocabulary based on the Exif and raw low-level data. This early prototype enabled us to realize:

- The hue, saturation and lightness of a photograph's dominant colors can provide a subjective expert vocabulary, hence provide the domain model.
- The injection of expert knowledge can be achieved through SARA [Hampson, 2009].
- The UI must be enhanced to provide more efficient and effective end-user exploration

These deduction and requirements were added into the design process in order to produce the final prototype. After the completion of the application, user-centric usability and acceptance testing took place to evaluate the effectiveness of the tool as a means to visual search and exploration.

## 1.5 Thesis Outline

The remainder of the thesis is organised as follows:

- Chapter Two presents the current state of the art in image retrieval, focusing on high-level semantic search systems.
- Chapter Three outlines the design of an application with particular detail on the features identified as a result of an initial mock-up analysis.
- Chapter Four details the implementation of a tool that empower users to explore large volumes of visual data using semantic criteria to refine and focus their needs.
- Chapter Five shows the results of the evaluation of the final tool referred to as $X^2Photo$.
- Finally, chapter Six presents the results of the research and the conclusions derived, as well as avenues of further research in the area while suggesting real-life applications of $X^2Photo$.

# Chapter 2

# State of the Art

## 2.1 Introduction

This chapter presents the current state of the art in indexing and retrieval of digital content, specifically photographs and overviews current exploration engines and tools. Section 2.2 describes the underlying problem, the so called semantic gap and presents an outline of the current basic techniques seeking to address it. Section 2.3 focuses on image search engines and a popular photograph sharing site based on both their capabilities and aesthetics. In section 2.4 image retrieval based on content is examined and applications of the field are viewed. Section 2.5 presents an overview of a knowledge discovery system and section 2.6 looks into color theory and its usage in the photography domain. Section 2.7, based on the review of the chapter, analyzes the current state in image retrieval and the possible improvements.

## 2.2 Attributing semantics to photographs

Most of the current research in image retrieval is connected with the semantic gap in image retrieval. In essence, the gap between the low-level physical features of the image and the high level perception of what the image portrays is referred to as the semantic gap:

*"The representations one can compute from raw image data cannot be readily transformed to high-level representations of the semantics that the images convey*

*and in which users typically prefer to articulate their queries."* [Hare et al., 2006]

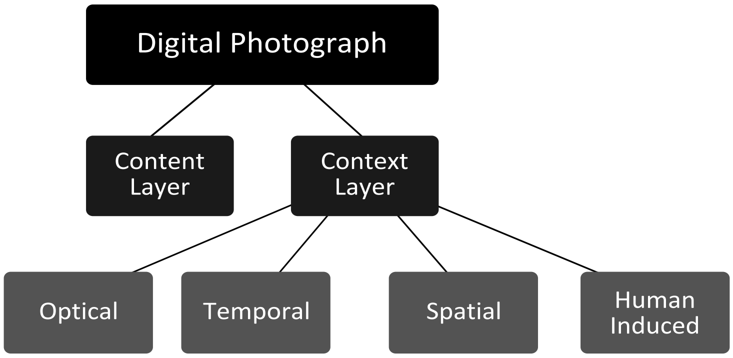



**Figure 2.1**: A digital photograph

[Sinha and Jain, 2008b] propose that a digital photograph is a multilayered structure comprising of a content layer and a context layer. The content layer stores data recorded by the CCD as pixel values. The Context Layer stores the contextual information about a photo shoot and is divided into the following four sublayers.

- **Optical Context Layer** contains the metadata related to the optics of the camera; e.g., the focal length, aperture, exposure time etc. These values define the context in which the image was shot (lighting condition, depth of field, subject distance).

- **Temporal Context Layer** contains the time stamp of the instant that the photo was shot.

- **Spatial Context Layer** contains the coordinates of the location the photo was shot.

- **Human Induced Context Layer** contains the tags created by people.

This classification is most helpful in analyzing the basic techniques currently used in image retrieval.

### 2.2.1 Content Analysis

This approach is based on the idea of analyzing the content layer of a photograph, i.e. it aims to extract meaning from the recorded pixels. Techniques such as face recognition and region classification are based on this analysis. Regarded as a low-level approach, its attempt to close the gap is questionable [Datta et al., 2005].

### 2.2.2 Context Analysis - Exchange Information

These techniques rely on the metadata set at the time of digital image capture. Exchangeable image file format for digital still cameras, Exif, is a standard which specifies the formats to be used for images, sound and tags in digital still cameras and in other systems handling the image and sound files recorded by digital still cameras [JEITA, 2002]. A photograph taken with a typical digital camera contains Exif metadata about when the photograph was taken while some cameras also record the location, as well as optical information such as exposure time, focal length and the use of flash. In the image retrieval field, this metadata that is independent of the scene content has been used along with content analysis to produce photograph classification [Boutell and Luo, 2004], [Tuffield et al., 2006].

### 2.2.3 Context Analysis - Human Input

Utilizing user support in image classification can be divided into two sub classes; manual association and image retrieval based on relevance feedback.

With the development of online photograph sharing sites such as Flickr and Picasa, and the popularity of the ESP game [von Ahn and Dabbish, 2004] [Gwap, 2009] annotating images with related keywords has become a widely used technique to handle image sharing and retrieval. However as the effectiveness of this approach relies on the efforts of individuals, systems that depend only on annotated data cannot realize the full potential of a photograph collection. The required data will not always be available for every image. Also relying on natural language parsing entails inaccurate results due to synonyms.

Relevance feedback systems involve users in the retrieval process by enabling them to determine if the results they receive for their queries meet their information needs; relevance is based on user needs rather than a query. An image is regarded to be relevant if it addresses the user's need and not just because it matches the certain query. As users interact with the system during the retrieval process by refining their queries, gradually the result set is improved to produce an improved final result. A system using relevance feedback in retrieving images, dynamically learns the user's purpose and progressively provides more relevant images. Basically, the idea is to allow users to perform keyword searches and then use the given feedback to expand the image search results based on visual features. This technique has been combined with both keyword and content analysis to produce better results [Uchihashi and Kanade, 2005], [Jing et al., 2005]. Although useful for generic searches, in the photography domain assuming user interest is only tied to the *content* of a photograph will cause inaccuracies. Users' focus might be on the general location rather than the object or maybe on the overall effect the photograph has on them.

## 2.3 Keyword-based semantic search

Image retrieval based on keyword features [Tamura and Yokoya, 1984], [Shen et al., 2000] was mainly developed by the database management and information retrieval community. The typical query scenario in such image retrieval systems is Query By Keyword (QBK). Semantics of images can be represented by keywords, if these keyword annotations are accurate and complete. But as the size of the image database gets larger, manual annotation cannot be regarded as a logical process. Popular image search engines assume to overcome this issue by extracting the keyword features surrounding an image on the Web. The problem here is that such automatically extracted keywords are far from being accurate and complete.

### 2.3.1 Image Search Engines

Web-based services that gather and index images from other sites on the Internet are referred to as image search engines. This service is offered by general search engines, such as [Google, 2009] or [Yahoo!, 2009] which are the major image search providers along with Microsoft's [Bing, 2009] which has recently become a significant competitor, third in terms of size and usage, as well as some search engines focusing entirely on images. [Picsearch, 2009] and [Pixsy, 2009] are among the notable specialized image search engines. Another class would be online catalogues such as [iStockPhoto, 2009] which is a royalty free, international micro stock photography provider whose parent is the [Images, 2009].

### Image indexing and retrieval

Online image search engines rely on the textual information surrounding or associated with an image for their indexing. The information within an HTML document; such as the page title, "alt" picture tags or the explanatory text surrounding the image as well as the URL and filename of the image itself, is used to index images in typical image search engines. Although this method can find numerous images, especially when you have an enormous index like Google's - many times that of Yahoo and Bing - the returned images are not entirely accurate since the surrounding textual information cannot guarantee that the image itself is in fact characterized by the associated keyword. For example when running a simple search with the keyword "stone", Bing returns an image of "The Long Stone", in Parish of Minchinhampton, Gloucestershire as the first result which is followed by Sharon Stone or Rolling Stone magazine covers as they all had the term stone in their filenames: EMLongStone1913.jpg, stone.jpg, rolling_stone.jpg respectively. This problem caused when trying to attach semantics to visual content is of homonymy, where the same tag may have various meanings [Golder and Huberman, 2005]. Hence engines that retrieve images indexed through such methods can only be accurate within a certain limit [Cai et al., 2004].

Another weakness of keyword-based engines is that when a query contains ambiguous and abstract search terms, as long as they are not directly associated with the image itself, it is impossible to obtain relevant results. Approaching images with direct textual keywords forces them to lose their expressive power and visual value. In order to overcome this issue Google chose to have people label visual content instead of a search engine parsing. In 2006 they licensed Luis von Ahn's ESP Game [von Ahn and Dabbish, 2004] [Gwap, 2009] and introduced the [Labeler, 2009] service. In this game two random users are asked to tag the same image simultaneously. If the given answers are the same, the application assumes that the label is in fact viable and stores it, otherwise it is ignored. Several drawbacks directly come to mind: How do two different individuals tag an image? Do they base their decision on the content or the overall impact? Also as the players are competing against time, would they actually try to be as precise as they can be or just point out the obvious?

[Golder and Huberman, 2005] answer that with the *Basic Level* problem. Terms related to an item can range from very general to very specific. But *"the basic level, as opposed to superordinate (more general) and subordinate (more specific) levels, is that which is most directly related to humans' interactions with them"*. In other words, players would label the image with the basic term. But then the intellectual levels of the individuals would cause a conflict. That is if one of the players is an expert on the subject, even in their most basic words, they can be more specific than the other player. A person with no domain expertise may just tag the image with "snake" while a snake expert would choose "python", describing it in the most basic level, but of their own.

Whether parsed automatically or tagged by humans, the search engine will still bring many images from several clusters that are mixed together and the user will have to carry out a time consuming process to find the image they are looking for. In order to enable the user to make more refined queries Google, Yahoo and Bing provide simple filters to set the size and related tags which aim to narrow the query by categorizing. Both Yahoo and Bing allow the user to choose whether they are searching for black and white or color images while Google provides a few colors to be selected in order to have results with similar ones. Also by using tags produced via content-recognition software processing, Google and Yahoo along with [Exalead, 2009] enable users to filter images by selecting the face choice.

Picsearch has considerably less amount of indexed images compared to general search engines. However, it allows its whole collection of images to be viewed where others cap their results around a thousand images. Picsearch provides similar filters to refine the search and having access to all their indexed images enables users to have a slightly higher chance of finding an accurate image, even though the process can be wearisome. Pixsy's difference comes from the fact that it obtains its results from selected RSS feeds, which has an upside of displaying users with images that are not in the scope of others and the downside of eventually providing a quite narrow result set.

As the above mentioned search engines pull their images from the Web it is not likely to find high quality photographs. An exception would be Yahoo as they have acquired [Flickr, 2009], a rather large photograph-sharing site which will be discussed in the next section. This acquisition makes Yahoo a preferred search engine when users wish to find good quality photographs as content can be pulled from both the Web and Flickr. iStockPhoto apart from having a wide range of stock photographs as well as vector illustrations and Flash files (also video footage and audio tracks), has an advanced search feature which significantly surpasses that of other web-based search engines. A notable filter is searching with color models which are; simple - much like Google's, web colors, RGB (Red Green Blue) and also HSV (Hue Saturation Value). The user can also choose to enter the hex value of a color.

Another unique feature iStockPhoto provides is the CopySpace. It is a 3x3 grid which aims to solve the layout issue. By clicking on the cells of the grid one can specify which areas they wish to be empty, occupied or unspecified. The result set then comprises images with subject and empty space based on the rendering of the chosen grid placement. While they provide high quality photographs with an aesthetic appeal, iStockPhoto is rather intended for designers and seekers of images to be utilized in specific cases such as ads, magazine articles or websites.

### User interfaces

In terms of usability, commercial search engines put only a slight focus on the user interface. Even though one might argue that their approach is towards a simple yet efficient design, the time and effort that a user needs to expend in order to explore the retrieved images doesn't indicate efficiency. The list based presentation that they utilize display images with different semantic concepts or visual features mixed together. Browsing page-by-page scanning dissimilar images prevents users from grasping the big picture, thus significantly diminishes the chances of carrying out context-aware consecutive searches.

[Broder, 2002] present a taxonomy of web searches and classify web queries according to their intent into three classes:

- **Navigational** where the intent is to reach a particular site.
- **Informational** where the intent is to acquire some information most likely from multiple sources.
- **Transactional** where the intent is to reach a site where further interaction will happen, such as shopping, finding various web-mediated services, downloading various type of files, etc.

[Hoeber and Yang, 2006] analyze the user information needs for web image search results browsing. Based on findings from [Broder, 2002], suggest that in navigational search activities the list based presentation of typical search engines may suffice. But these interfaces do not support informational search activities, in which a user wants to find out about a specific topic or get an answer to an open-ended matter. As the ranking algorithms aren't adequate enough, the images displayed on the first couple of pages may not be necessarily better than the following ones and the user needs to explore one set of images after another. Furthermore, as users cannot gather similar images together, they cannot perform comparisons between images.

To provide an efficient exploration [Hoeber and Yang, 2006] propose a similarity based search result presentation in which the retrieved images are re-organized and presented to users based on either image concept or image appearance gathered. The retrieved images are embedded in a two-dimensional layout which may cause overlapping of images depending on the number of results, leading to some images or parts of them to be invisible to the user. They attempt to overcome this issue by presenting a slider bar which let users adjust the spatial position of images to modify the overlapping ratio. This ratio can be adjusted along the range $[0, 1]$ where $\gamma = 0$ being the most overlapping layout and $\gamma = 1$ generates a fit-to grid view. Their experimental evaluations show that the similarity-based layout speeds up the overall search process.

[Villa et al., 2008] aim for a faceted retrieval interface which allows multiple searches to be executed and viewed simultaneously by splitting the view into multiple panels. Each panel represents a single facet of a larger task. With transitions via drag and drops between panels material can be reorganized between the facets. The user is enabled to carry on multiple related searches simultaneously. Another attempt to present users with several panels is presented in [Xu et al., 2009]. When a user submits a query the interface retrieves the results and organizes them in separate panels. The idea is to detect the topics that are latent in the result set and then classify the search results into clusters by their topics. Both studies conclude that their visual search interface, enable users to answer explorative questions with less time and mouse clicks.

Instead of displaying images in a two-dimensional layout, the web browser plug-in [Cooliris, 2009] provides interactive picture and video exploration over a three-dimensional interface. The tool enables users to search and view images from various sources such as Google Images and Flickr. The 3D Wall embedded in the interface displays the visual content in a cinematic way. The significance of the wall-like display is that it enables users to get a clear view of the retrieved images at a single glance. Users can adjust the zoom in order to observe single or multiple images. Unlike the list-based presentation, the users have a smoother browsing experience as they do not need to reload new pages to navigate. Its three-dimensional nature allows large amount of images to be visible at once without any overlaps as opposed to flat layouts.

Even though image search engines are commonly used when users wish to find general images, they cannot be regarded as sole providers. Limitations brought on by discussed characteristics prevent them from providing accurate and precisely relevant results which have good aesthetic qualities.

### 2.3.2 Flickr: A photograph-sharing site

Flickr, an online community platform, enables its users to upload, store and organize digital photos, as well as to automatically post camera-phone shots to a blog. On

November 2008, they announced their "$3billion^{th}$!" photograph [Champ, 2009], and as of June 2009, are estimated to host more than 3.6 billion photographs from over 90 million users [Mislove et al., 2008], holding a noteworthy place among social media sites.

### Organization of photographs

Flickr's content organization methods not only help users define and categorize their collections but also keep adding to the descriptive metadata associated with the content.

- **Setting titles and image descriptions** Help users to organize their individual photographs and express the photographs in a personalized way.

- **Tagging** The tags in Flickr are mostly assigned by the owners [Marlow et al., 2006], hence they do not support collaborative tagging [Golder and Huberman, 2005]. This aspect suggests that photographs within Flickr will be surrounded by more subjective concepts. Assuming that a user sharing a photograph is trying to get a message across, this notion should also be evident in the tags they choose. Although this cannot be generalized for everyone, the increasing number of tagged photographs provide indication that users accept tagging as a means to express their content on Flickr [Marlow et al., 2006].

  [Nov et al., 2008] investigate the various factors that people consider when tagging their photographs on Flickr by using survey results from [Ames and Naaman, 2007] and enhancing it with Flickr system data about actual usage. Their research suggest that users tend to tag their photographs in order to create a social presence among their contacts as well as the general public rather than organizing them, supporting the idea that tags within the Flickr collection are much more relevant and have a stronger power of expression associated with them.

- **Classifying** By allowing users to group their photos into sets, and their sets in to collections, and so forth, Flickr wraps the object with a descriptive categorical metadata. Hence the object becomes more aware of itself when it comes to retrieval.

- **Adding images to groups** Flickr allows users to create special interest groups on any possible topic. Like classifying, the groups that a photograph belong to strengthens it available metadata.

### Exploration

The great difference between Flickr and commercial search engines lies in the exploration of images. Constantly making use of the metadata that surrounds a photograph, which is significantly populated via the "organization" methods, Flickr guides the user through the actual exploration. Searches can be made based on full text and/or tags as well as through group pages. The Explore page allows different views based on time, geographic location, popular tags, etc. A Flickr feature called Interestingness allows the browsing of various photographs.

*"There are lots of elements that make something 'interesting' (or not) on Flickr. Where the clickthroughs are coming from; who comments on it and when; who marks it as a favorite; its tags and many more things which are constantly changing. Interestingness changes over time, as more and more fantastic content and stories are added to Flickr."* [Flickr, 2009]

Within Flickr, users and their contacts form the backbone of photograph propagation. Research indicates that social browsing, i.e. finding photographs by browsing through the photograph streams of contacts, is one of the primary methods by which users find new images on Flickr [Lerman and Jones, 2006]. This brings up an interesting aspect; that in such an environment users are likely to "follow" the likes of others to explore visual content suggesting that photograph enthusiasts welcome the idea of expert guided browsing.

## 2.4 View-based semantic search

Research activity in visual image retrieval increased following the adoption of a new approach: Content-based image retrieval (CBIR). CBIR is the method of retrieving images on the basis of automatically-derived features such as color, texture and shape. These systems try to retrieve images that are similar to a specification or pattern (e.g., shape sketch, example image) a user defines. The automatic retrieval process within these systems suggest an advantage compared to keyword-based search systems as there is no possibility of the necessary metadata not being present.

IBM's QBIC [Flickner et al., 1995] is a commercial pioneer in CBIR. QBIC allows queries on large digital media databases based on example images, user drawings, selected color and texture patterns and other graphical information. [Hermitage, 2009] Web site uses the QBIC engine for searching archives of world-famous art. Query By Visual Example (QBVE) has the shortcoming of not understating the appearance of an image but rather just finding similar images. Therefore, this method does not address the semantic gap, as users can not specify their queries through natural language descriptions.

In order to overcome this matter, Query By Semantic Example (QBSE) [Rasiwasia et al., 2007] was developed. This method can be described as QBVE with Semantic Retrieval (SR). In this approach, models of image concepts are created that capture the relationships between words and images. SR enables high level queries while QBVE tackles the problem of lexical ambiguity. The system learns what a "sky" is from other images and starts associating the word "sky" with similar images it finds through the QBVE method. This method has the ability to perform semantic deduction, connect the appearances of different images to their labels in order to form global image topics, but still the system cannot provide a subjective aesthetic perspective in image classification.

Some widely used commercial visual search solutions for the stock photography industry are Idée's [Piximilar, 2009] and Similar Image Search from [imense, 2009]. Photology [Enotic, 2009] is a software that enables users to browse and organize their collections based on filters such as time, indoors/outdoors, photo content (plants, sky, faces, beach, flowers, snow, sunset, water) and color, without using any tags. Filters are combined to submit refined searches, but as these filters are increased or if they conflict (indoor photograph of a beach) the result set is clearly irrelevant both within and with the query.

Comprehensive surveys on content-based image retrieval can be found in [Datta et al., 2005], [Lew et al., 2006] and [Smeulders et al., 2000]. The limitations of current content-based retrieval approaches and their incompatibility between searchers' queries are often pointed out [Hare et al., 2006] [Enser et al., 2006] [Eakins et al., 2004]. The major obstacle in content-based image retrieval approaches is the gap between visual feature representations and semantic concepts of images.

In general, the problem with these algorithms is their dependency on visual similarity in judging semantic similarity [Datta et al., 2005]. Especially for photographs, it is very difficult to devise effective features that reflect their aesthetic characteristic. As semantic similarity is a highly subjective measure, it is not reasonable to rely on such algorithms, especially when the semantic space comprises of aesthetic values.

## 2.5 Expert assisted knowledge discovery

Knowledge discovery can be considered as the extraction of new and useful information about an application domain, in some sense deriving knowledge from the data [Frawley et al., 1992]. While Knowledge Discovery in Databases (KDD) involve applying knowledge discovery processes to databases, the field of Knowledge Discovery and Data Mining (KDDM) is concerned with applying knowledge discovery processes to any data source [Kurgan and Musilek, 2006]. There are several different KDDM process models that consist of multiple steps executed consecutively. In [Kurgan and Musilek, 2006] a generic process model that consists of six steps is specified in order to provide a common framework.

Semantic Attribute Reconciliation Architecture (SARA), a framework that specifically addresses the need for tools that support the user-centered exploration of heterogeneous information sources, developed in the Knowledge and Data Engineering Group [KDEG, 2009] at Trinity College Dublin, allows users to construct personalized and semantically meaningful queries, which leverage expert knowledge to assist a user in exploration of information [Hampson, 2009].

The underlying methodology tailors the generic process model [Kurgan and Musilek, 2006] specified by introducing personalized querying in the fourth step in order to bring in expert knowledge into the system as well as personalization. SARA seeks to encode expert knowledge to support novice users' involvement in semantically enriched and personalized explorations of heterogeneous information sources in order to enable the organization and management of large volumes of diverse data encountered in everyday life. In this sense, SARA can be considered as a powerful semantic intermediary between end-users and the raw information sources they seek to explore as it allows end-users to employ expert knowledge as semantic attributes to create high-level queries. As it allows users to extract additional meanings and create relationships between them by tying them via complex queries, SARA injects "intelligence" necessary for such systems that focus on information retrieval.

## 2.6 Color in photography

Within CBIR approaches a good amount of attention has been focused on color, as a key feature to characterize the content of digital content collections [Gong, 1999], [Jau-Ling and Ling-Hwei, 2002], [Yu et al., 2002]. Colors are defined on a selected color space and different color spaces are utilized in a range of applications. [Plataniotis and Venetsanopoulos, 2000] gives a description of various color spaces; widely used are RGB, LAB, LUV, HSV (HSL) and YCrCb. Common color features include, color-covariance matrix, color histogram, color moments, and color coherence vector. Even though these color features are efficient in describing colors, they are not directly related to high-level semantics.

One way to derive human perception through colors is to investigate the psychology of color in art [Davis, 2000], [Gage, 1999]. Artists use color to explore visual perception and to represent or evoke emotions. The psychological effects of color hue, saturation, and brightness have been studied to reveal having various effects on the viewer [Fehrman and Fehrman, 2000], [Mehrabian and Valdez, 1994]. Although the concept of color symbolism is an unstable cultural construct, varying with time and place, a “transcultural” approach has been suggested in [Morton, 2004]. According to [Morton, 2004] nature provides a starting point for universal color symbolism and that natural references can be considered timeless and cross-cultural. Other symbolic meanings linked to politics, gender, age, etc may change and are considered era specific.

Color theory is a language that conceptually and perceptually describes the essentials of color and their interactions [Parramon, 1989]. It is impacted by color management (the science of measuring color physically to describe it mathematically), color adjustment (techniques for altering colors) and color psychology.

Unlike color psychology, color theory doesn’t describe responses that are unique to cultures or certain periods, but rather focuses on universal psychological responses to color. An example would be the warmth or coolness of a color, i.e. the temperature. Colors such as blue and green are cool colors where red and orange are warm colors. Cool colors can be thought of as having calming effects while warm colors are perceived as exciting. This effect can change as the color’s luminosity changes, i.e. a sunset can be regarded as soothing while a bright open sky may be exciting. The warmth of red, yellow, or orange can evoke enthusiasm or anger. Warm colors express emotions from light optimism to strong violence. Similarly cool colors on one end can be cold, impersonal, and gloomy but on the other comforting and nurturing.

In photography color theory is utilized to understand how certain colors and their combinations create different moods in photographs. Some color combinations such as complementary colors appear striking and vibrant when in close proximity. Also as they get closer to the same saturation and lightness, the vibrant look will strengthen. On the other hand, colors that are close in the spectrum will usually appear more peaceful and calm.

## 2.7 Analysis

Looking at the current approaches, it is apparent that bridging the semantic gap is still an open issue. Indexing based on surrounding textual information is highly unreliable, and textual annotations depend on the knowledge and expressiveness of individuals which causes ambiguity. Retrieving images through these textual data result in inaccurate and irrelevant clusters of images because of the system's blindness. Also the limited capabilities of tagging prevent the aesthetic appreciation of quality photographs. Current low-level visual information based retrieval technologies do not allow users to search for images by high-level semantics. The need to provide initial query images or to find images based on unintuitive low-level characteristics clearly explains why these approaches haven't yet found a noticeable place in the commercial world. Regarding photography appreciation both approaches though acceptable for defining content, are inefficient in reflecting the aesthetic characteristics of images.

[Sinha and Jain, 2008b] point out that content only is not enough in inferring the semantics of photographs. They suggest fusing content and context to extract semantics; referred to as a *Contenxtual Analysis*. [Enser, 2000] also suggest that it is necessary to utilize both the concept and the content of a photograph to improve the efficiency of image retrieval techniques stating that hybrid image retrieval systems should be welcomed.

To make the most of both keyword-based and view-based approaches' strengths in image retrieval, various approaches have been suggested to combine these features. [Addis et al., 2003], [Lewis et al., 2004], [Boutell and Luo, 2004] and [Sinha and Jain, 2008a] suggest that systems integrating image content and camera metadata improve the image retrieval process. The tendency towards combining keyword and view-based search can also be seen through the recent features of popular Web search engines, like that of Google, Yahoo, Exalead and Bing, where options like choosing color and searching for portraits are being provided. Pixolu [Pixolu, 2009] is one of the modern prototype visual image search softwares that combine keyword search, visual sorting and visual similarity search, and semi-automatically learned semantic relationships between images.

When focusing on image retrieval with subjective qualities such as aesthetics it is useful to have image retrieval systems that can introduce personalization into the whole process while facilitating intelligent queries. Being able to retrieve images from sparse sources using high level semantic attributes will enable these systems to fully realize their potential. The language of color theory presents a semantic space that can be associated to low-level features of photographs to create a perceptive and aesthetic approach towards searching for good quality photographs. Combined with a user interface that support end-users in manipulating these photograph collections in a personalizable and compelling way, such a system would empower users in exploring and actually accessing the million shots.

# Chapter 3

# Design

## 3.1 Introduction

This chapter presents the design decisions considered in order to develop the tool which will empower users to explore large volumes of visual data using semantic criteria to refine and focus their needs. Section 3.2 analyzes and details the key requirements for such a tool. In section 3.3 the overall architecture is detailed focusing on each component. Section 3.4 focuses on the exploration space with respect to individual units of interaction and section 3.5 provides a summary of the chapter.

## 3.2 Requirements

### 3.2.1 Analysis

Related work showed that image retrieval techniques are tending towards multi-modal systems. Specifically, approaches combining textual annotations/keyword search and low-level characteristics seem to be considered in order to bridge the semantic gap, thus provide more efficient and effective systems. However, we cannot conclude an ideal content-based image retrieval technique, as various systems apply various algorithms relying on sometimes shape, sometimes regional shape and color, texture or different combinations, etc. We have emphasized the need to facilitate aesthetic appreciation of photographs. In order to realize this, our approach should be able to offer aesthetic means rather than just the content itself and enable the end-user to extract this knowledge in a natural and intuitive manner. Indulging in the fact that this aesthetic value is highly subjective, end-users should be provided with a flexible semantic space. End-users should also be aided in the process of exploring the system's sources. The retrieval of images should not be considered as a simple send-query-receive-result process, but rather as an interactive and engaging exploration seeking to involve the end-user in the refining decisions.

Based on this analysis, we derive a number of key requirements considering several aspects as discussed next in this section.

### 3.2.2 Manipulation of Tags

The single dimensional nature of tags has been pointed out discussing that they are limited in communicating the aesthetic values of photographs, but are instead useful in defining their content. In this research we are investigating if the "picture" and the "photograph" can be retrieved. Using objective human input will allow us to investigate the picture and as tagging has become a widely used approach in order to express content, it will be required to integrate into the retrieval process. Therefore the system should enable end-users to view the tags associated with photograph collections obtained from query results as well as each individual photograph in order to define the labels for consecutive searches. The manipulation of tags must be decided upon during the testing of a mock-up version of the tool.

### 3.2.3 Color Theory Integration

CBIR methods are considered when characterizing the low-level content layer of collections. By analyzing the automatically-derived elements, we can gather the visual feature representations which will later be associated with semantic concepts. In order to form the low-level layer we will focus on color. Based on related work we derive that for representing color, the HSL color model will be most efficient regarding this research's aims.

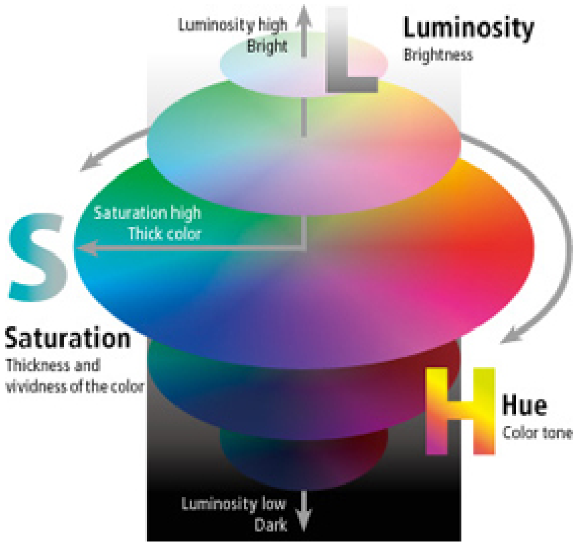



**Figure 3.1**: The HSL color space can be represented by a double cone showing the three axes of hue, saturation and luminosity. [Canon, 2009]

It has been observed that RGB colors have limitations such as being hardware oriented and non-intuitive [W3C, 2009]. HSL color space describes perceptual color relationships more accurately than RGB and is far more intuitive. The HSL color space defines colors more naturally: The *hue* of a color is determined by its wavelength and is represented as an angle of the color circle. The terms "red" and "blue" are primarily describing hue. Simply put, it is the term that we basically refer to when describing a color. *Saturation* is the intensity of a particular color; 100% is full saturation, unsaturated color tends towards grey, and 0% is a shade of grey. 0% *lightness* is black, 100% lightness is white, and 50% lightness is 'normal'. HSL color space is most efficient because it is closely related to human visual perception. Color quantization is also useful for reducing the calculation cost and provides better performance for semantic ruling as it can eliminate the detailed color components which can be regarded as noise. Another point deducted from color theory is that the human eye is more sensitive to hue than saturation and lightness. Therefore hue should be processed with a finer quantization.

### 3.2.4 Expert Vocabulary

An expert vocabulary based on the manipulation of raw low-level data of digital photographs should be created and embedded into the system. This vocabulary will be an extension of the language that color theory provides. Subjective concepts will be introduced integrating color psychology. A photograph's dominant colors' HSL space will indicate the photographs subjective aesthetic quality. The system should maintain a flexible underlay as the introduced vocabulary will be highly subjective. The vocabulary will enable the end-user to access the photographs from an aesthetic perspective and this perspective will depend on the expert's view. The user may adapt to this perspective or may find it open to questioning. Therefore, such flexibility must be preserved to have a subjective system. The important factor here will not be the vocabulary, but providing a base that a subjective concept can be built upon will.

#### Injecting Expert Knowledge

The system requires a framework that will act as a semantic intermediary between end-users and the raw information sources. As mentioned, we must obtain a flexible base while using a subjective concept. In order to provide such an extensible system, the intermediary framework must be flexible while associating highly subjective concepts to low-level definite characteristics. Based on this necessity, this research will utilize SARA [Hampson, 2009]. As investigated, the chosen framework enables end-users to easily create new semantic spaces from the extracted raw low-level features. Another reason behind this decision was that SARA enables users to construct personalized and semantically meaningful and complex queries, which is necessary in this research as we seek to provide full access to a photograph's aesthetic qualities; the end-user must be able to refine their search via consolidated aspects. While integrating SARA into our system, optimization and efficiency must be taken into account. The design should be produced regarding this factor.

### 3.2.5 User Interface

This research seeks to build a tool with a highly visual user interface. This system must focus on the presentation while sustaining high usability and functionality. Also the limitations of current applications discussed in the previous chapter must be considered when developing the user interface.

#### Interaction

The system needs to provide several access points, i.e. the tag space, subjective expert vocabulary and the discovery space as well as refining, during the exploration process. Each element needs to be easily interactive and their collective usability must also be considered. Their presentation and position in the interface must be flexible to provide the end-user with effective control and freedom. Considering an approach that will enable the end-user to decide when and where to display these elements will give them a feeling of both power and ease, hence improving the interaction quality. The end-user's interaction with individual photographs must be informative and visually pleasing. They must be able to view the original photograph as well as a thumbnail version, while not breaking the exploration and sustaining a smooth transition. Overall the system should provide an interface that has a minimal learning curve and it should increase accuracy and efficiency of the exploration without losing usefulness, and must not be frustrating and confusing to carry out a task.

#### Personalizable Refining

One of the key requirements of this system is to introduce the concept of refining a search focusing on a particular photograph. End-users must be able to carry out consecutive searches based on a photograph that has similar features with the photograph they seek to find. By adjusting the aesthetic quality of the photograph in focus via interacting with the tag space and the expert vocabulary, they will be able to refine a search. This aspect will enable them to express their target more efficiently and gain an idea of the expert's perspective. As a photograph is brought into focus, both its tag space and aesthetic qualities must be displayed to guide the end-user. Through this approach the system will empower end-users to carry out personalizable queries.

### UI Influences from SoA

Within photograph collections, sometimes even with individual photographs, the number of crowd sourced tags exceeds a number that can be clearly displayed with classic list-based presentations referred to as simple tag clouds. In order to overcome this issue, the system requires a representation which will provide access to extensive amount of related tags without crowding the display. The SoA indicated towards an approach where a tag cloud having tags distributed among the faces of a sphere would be most efficient. WP-Cumulus is a Wordpress plug-in [WP-Cumulus, 2009] which displays the tags in a sphere. It takes less space than the traditional tag cloud. It also gives a better aesthetic look than that of long list of tag clouds or a long list of tags. The tag space referred to as a *TagBall* in this research builds on the mentioned idea and improves it by adding dynamic resizing and drag-and-drop features which will be further discussed in the implementation chapter.

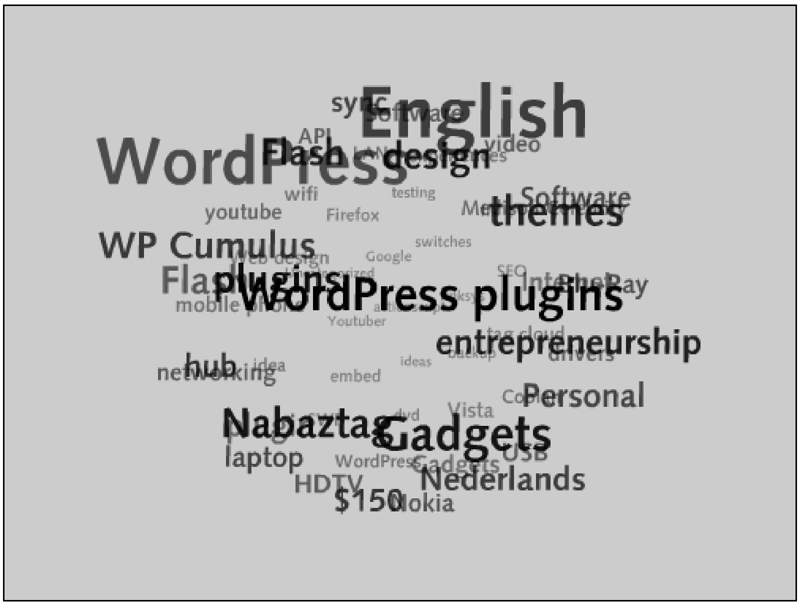



**Figure 3.2**: WP-Cumulus' tag sphere

Another necessary interface feature is the display of photograph collections in a manner that will support informational search activities. A combination of the similarity based search result presentation [Hoeber and Yang, 2006] with a three-dimensional wall view similar to that of [Cooliris, 2009] was decided to be the most efficient model. Gathering similar images based on different criteria catches the focus of the end-user towards the set they are targeting. The flow-like display of photographs gives the user a continuous browsing experience and the three-dimensional display improves both the usability as it maximizes the exploration space, and visual appeal. This research while adopting the feel of mentioned approaches does not in any way make use of the tools or plug-ins. Instead the concept was built from scratch with different technologies which is described in the implementation chapter.


**Figure 3.3**: Cooliris screen shot

## 3.3 Architecture

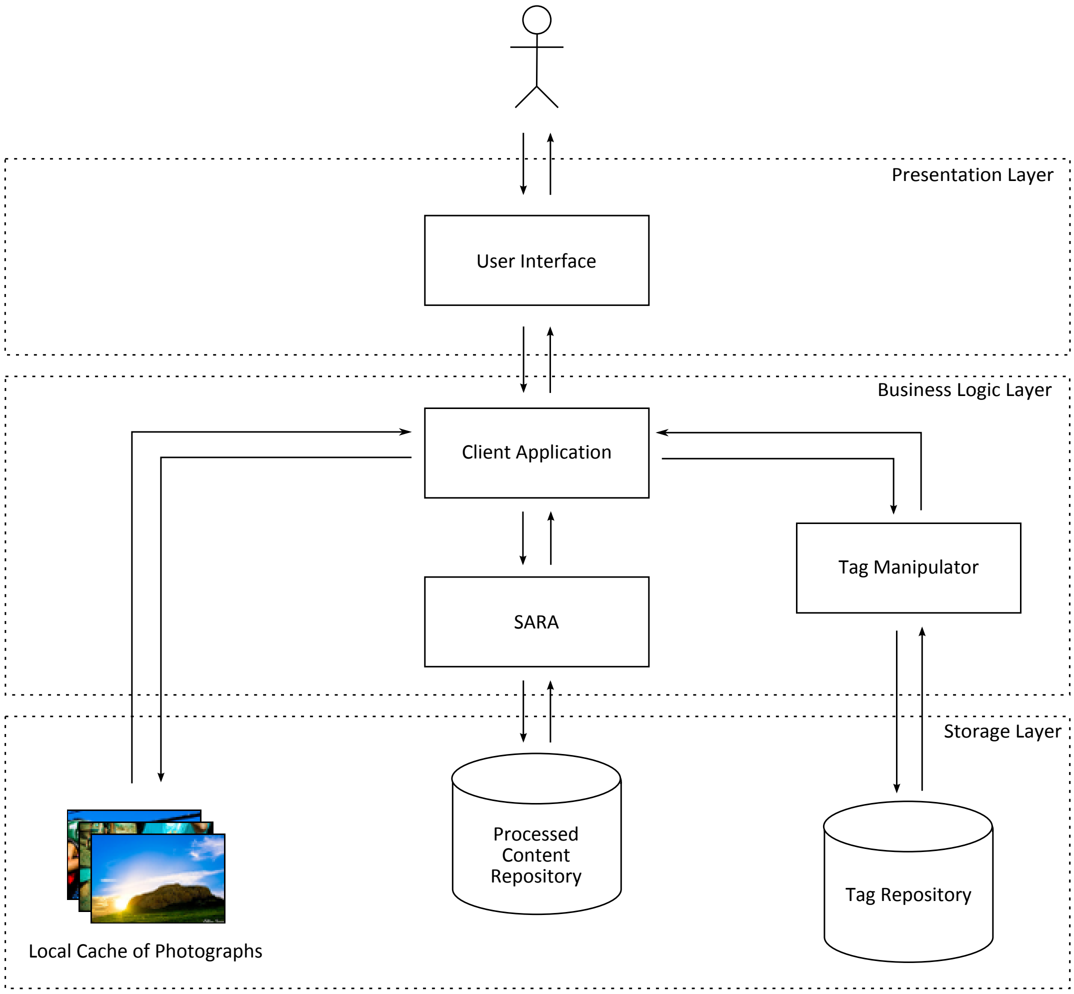



**Figure 3.4**: High Level Architecture

Figure 3.4 shows the high level architecture of the application. In this section individual components grouped within each layer will be detailed, and then their interaction will be presented as the system flow.

### 3.3.1 Presentation Layer

The presentation layer holds the user interface which enables the end-users to interact with the system. It will be thoroughly discussed in section 3.4.

### 3.3.2 Business Logic Layer

This layer comprises of the client application, SARA and the Tag Manipulator:

#### Client Application

The client application will be responsible for the data flow between the user interface and local cached photographs, SARA and the Tag Manipulator. Once initialized, it will configure the user interface and then start listening for events from it. The input received from the user interface will be the selected tags and the chosen semantic attributes that represent the expert vocabulary. The application will handle the user input through two channels based on the type. The chosen tag(s) will be communicated to the remote Tag Manipulator which will query the Tag Repository to acquire the related tags stored within. The semantic attributes, along with the received set of related tags, will be sent to SARA by calling the provided function through its API. The data received will be the identifier(s) of the photograph(s) that match the user query. In order to refresh the interface content, first the popular tags will be extracted and then the local cache will be accessed to retrieve the necessary photographs. Once the interface is refreshed, the client application will again go to listening mode.

#### SARA

In this design SARA will enable end-users to create semantically meaningful queries that carry the expert knowledge and related tags. It will be talking to the client application through its API. Once it receives the semantic attributes from the client, the system will depend on SARA to retrieve the matching photographs' identifiers and communicate them to the client application which will be listening for the results. The internal process model of SARA is beyond the scope of this thesis and can be found in [Hampson, 2009].

#### Tag Manipulator

In essence the Tag Manipulator operates as a service that facilitates the access of the client application to the processed set of tags retrieved from the repository. This remote service upon receiving the user selected tag(s) will query the related tag(s) from the repository and return the result set to the listening client application.

### 3.3.3 Storage Layer

The cached photographs, processed content and tag repository form the storage layer:

#### Local Cache of Photographs

One decision that was necessary to make was whether to access the photographs at run-time from the Web (i.e. a photograph sharing site that provides an API to retrieve photographs from its collection) or create a local cache. The reason that this question even arose was due to the fact that the initial approach was to make use of the Exif information of the photograph when creating the semantic attributes. After testing the mock-up it was decided that the color space would be used instead. Therefore, the latter approach of creating a local cache was chosen as the photographs needed to be processed in order to extract their low-level features. The local caching was also necessary for a smooth user experience.

#### Processed Content Repository

This repository contains the extracted low-level data of each photograph. SARA requires that a domain model be constructed and that the static sources containing the raw low-level data be stored in an eXist database [eXistDB, 2009] as XML files. Hence, this repository is an XML database containing the photographs' dominant colors' HSL space and also some information such as the photograph's title, owner, tags, etc. The domain model will be presented in the implementation chapter.

#### Tag Repository

During the implementation of the mock-up a MySQL database was created which contained the extracted low-level data and the relational table of tags. As the final implementation required an immigration to eXist DB, the tag repository was also converted into an XML file to be stored in the eXistDB. This file contains the tag clusters for each tag that was extracted from the whole photograph collection. This process will be described in the implementation chapter.

### 3.3.4 System Flow

The system is initialized through the client application which will load the user interface. End-users will be interacting with the system through this interface. They will run a query choosing the related semantic attributes and tags which describe their target result. The selected data will be passed on to the client application that is listening for events from the user interface. If any tag is received the application will first call the remote service of the Tag Manipulator which will query the repository to retrieve the related tags. This set will be returned to the client application which will then invoke the provided function by SARA passing the semantic attributes and the related tags. After SARA processes this data and returns the relevant photograph identifiers the client application will access the local cache to retrieve the necessary photographs and also gather the popular tags of the collection. This content will be made available to the user interface for display. The end-user will then continue to interact with the system in a similar fashion while the components will be in listening mode.

## 3.4 UI for Exploration

The system requires a highly visual explorative front-end which will empower end-users in manipulating large photograph collections in an engaging way, and enable them to interact with these photographs in a perceptive and authentic manner. In this section the spaces with which the end-user will interact is introduced and the possible use case scenarios are presented.

### 3.4.1 Tag Space

As mentioned, the main purpose of the tag space is to enable the end-user to define the content of the photograph, but of course as any tag can be chosen the system doesn't make a distinct separation. The tag space has two different states:

- Displays the popular tags of a photograph collection which can either be the whole collection in the case where the application is initially loaded or the tag collection of a result set obtained by a consecutive search
- Displays the whole tag set of each individual photograph

These states depend on the users' exploration view which will be discussed later in the Discovery Space section. In this design the tag space is referred to as the *TagBall*. The light three-dimensional geometry of the TagBall will enable resizing and also relocation by drag-and-drop. Having such a dynamic disposition will allow end-users to configure the interface to their like and preference, hence sustain ease and usability.

### 3.4.2 Expert Knowledge

The end-user needs to be extensive but also specific when describing the aesthetic qualities of the desired photograph when interacting with the system. Therefore an expert vocabulary which will satisfy this precision is necessary. The expert vocabulary's integration into the front-end will be achieved by the interface component that is referred to as the *AttBar* within this research. The AttBar will comprise of multiple sets of semantic attributes. Each set will present a semantic expression and its extended terms will be displayed on each bar. For example, if one of the main semantic attributes is "Temperature", its different levels will be displayed on the relevant bar, i.e. a bar with "Hot', "Warm", "Cool" and "Cold". Each of these attributes will be based on expert knowledge. A single bar can be thought of as a slider, where the user adjusts the level of the main attribute. Instead of having the user make adjustments with quantitative pointers (temperature-up-20, temperature-down-10) we chose to extend the attribute with natural language in order to provide a meaningful adjustment of semantic attributes. This way the user will be more expressive and intuitive and focus on quality, rather than thinking in quantities.

The display of the AttBar should fit nicely into the user interface and have a similar dynamism as the TagBall. End-user should be able to decide if the AttBar will be visible or not, expanding the Discovery Space and simplifying the whole front-end. Also rather than having a "plastic" look, i.e. a simple drop-box or a radio button like appearance, the AttBar should smoothly adapt to the overall feel of the interface. Based on the mock-up and several other design experiments the AttBar was shaped into its final appearance.

### 3.4.3 Discovery Space

The key element of the user interface is the Discovery Space which contains the photograph collection. Many display designs were considered throughout this research. The main considerations were:

- enabling an extensive view of the photographs,
- maximizing user interaction both with the collection and individual photographs,
- preventing the end-user to click any sort of buttons to refresh the view in order to display the next batch of photographs hence providing a smooth browsing experience,
- having the zoom functionality in order to display enlarged versions of photographs and their details,
- presenting the focus concept without any breaks,
- allowing a soft transition between the two views: Exploration and Favorites,

while maintaining high functionality, usability and a sleek look.

Before achieving the final Discovery Space different designs were considered to address these necessities. They will be discussed later in the implementation chapter. Also during the implementation of the final version, the front-end was tweaked constantly to attain an attractive interface while sustaining usability.
Taking all the requirements into account the final Discovery Space was designed accordingly:

- The Discovery Space is presented as a three-dimensional panoramic view which displays the whole result set.
- End-users can browse through the result set with simple dragging and can interact with individual photographs by simply grabbing them towards or pushing them back.
- As the whole result set is displayed the end-user can browse through the photographs in a flow-like manner.
- Individual photographs can be grabbed and viewed in zoom, and they can be flipped to access their information such as title, owner, camera model etc.
- Clicking on a button on the photograph enables it to be brought into focus which in terms of the UI mean pushing back rest of the collection and pulling the particular photograph on the z-axis, and enlarging it.
- The end-user can simply change view, while freezing the other, by sliding the panel back and forth.

Next we examine the possible use case scenarios in order to present the functionality of the front-end.

### 3.4.4 Use Cases

Figure 3.5 highlights the user interaction through different scenarios.

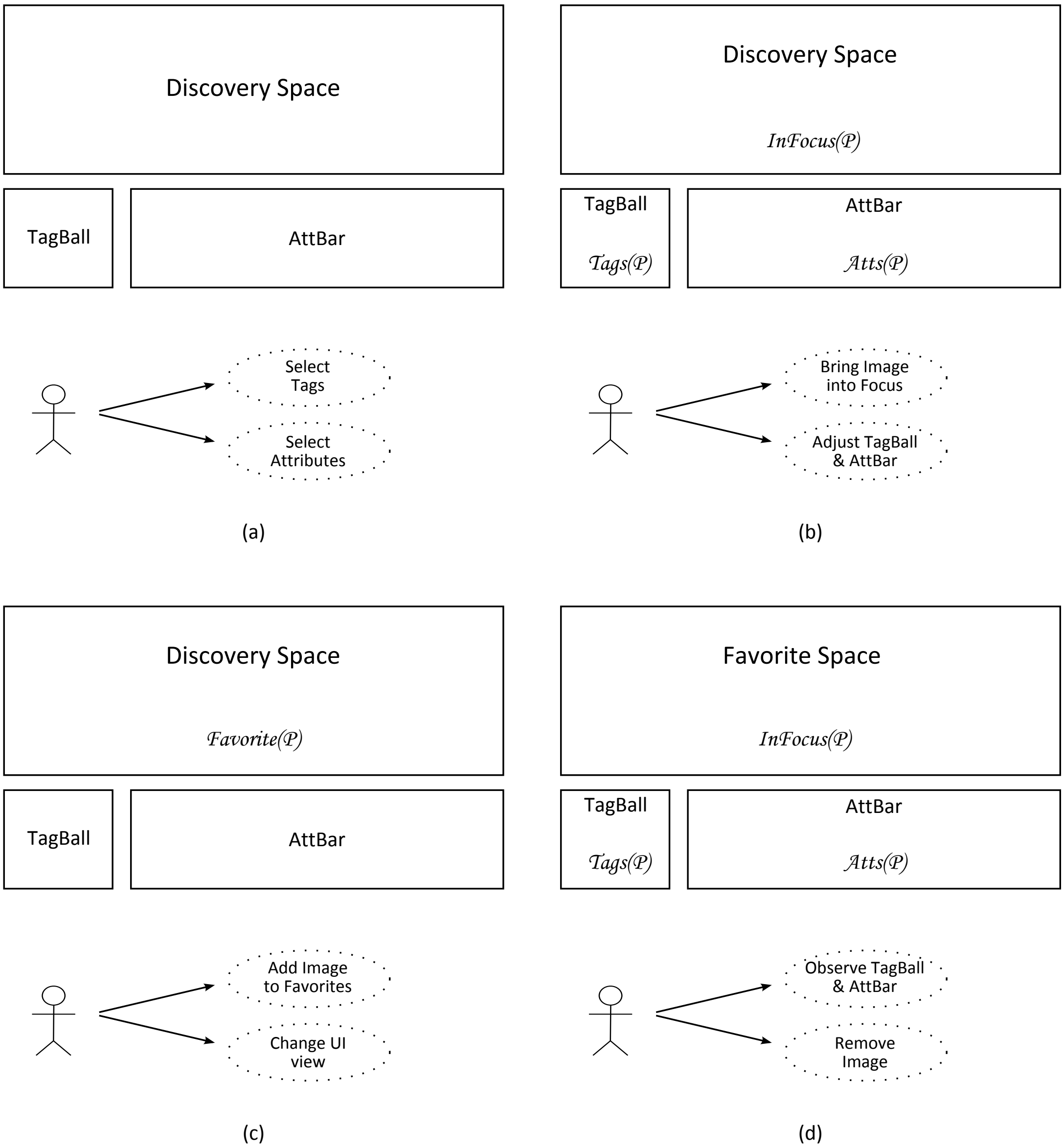



**Figure 3.5**: Use Case Scenarios

(a) In this scenario the user is either in the initial stage or has submitted a query and the view has been refreshed with the returned content. In any case, the end-user can enable the TagBall and select the necessary tags and enable the AttBar to choose the relevant semantic attributes in order to submit a new query. Also, if the user is running a consecutive search, a photograph can be brought into focus which produces scenario ($b$).

(b) Once a particular photograph $P$ has been brought into focus, i.e. $InFocus(P)$, the TagBall and the AttBar will be configured with the relevant content and display views $Tags(P)$ and $Atts(P)$ respectively. At this stage the end-user can carry out a consecutive query by adjusting these values (or keep them as they are if different from the prior search and match the required criteria). This functionality can also be used in order to recognize the expert's perspective.

(c) If a certain photograph $P$ is found to be suiting or just favored for some reason - $Favorite(P)$, it can be added to the Favorites Space. The end-user can accumulate such photographs in this space and access them throughout the session. The end-user can go back and forth between the two views by simply sliding the general container.

(d) Apart from providing a basket concept, the Favorites Space is also useful for the end-users to have an understanding of their own appreciation. The TagBall in this area is populated with the tags gathered from all the favorite photographs. If a user brings a photograph into focus in this area, the TagBall and the AttBar will again be populated according to the particular photograph's content. This space maintains a less condense discovery space with larger photographs. Also, when a particular photograph is zoomed into, the user can view it in its original size.

## 3.5 Summary

In this chapter the system requirements were presented, along with a textual and diagrammatical description of how the implemented system will meet them. A detailed description of the core components was discussed, including their roles and the interaction between. The explorative front-end was presented, introducing the *TagBall* and the *AttBar*. Also the Discovery Space and its functionalities were described, bringing in the concept of focus image based refining. Finally, possible use case scenarios were offered to clarify the functionality of the front-end.

# Chapter 4

# Implementation

## 4.1 Introduction

This chapter describes the implementation of the various components presented in Chapter 3 forming the system. The process undergone to finalize the complete system will be described in a chronological order. First, section 4.2 describes the data collection and the parsing process that was necessary to build the repositories. In section 4.3 the initial approach is outlined detailing the construction of an early prototype. Later in section 4.4 the various GUI prototypes that were considered are provided and the decision making behind each are given. The complete system is detailed in section 4.5 along with a demonstrative exploration and system analysis. Finally section 4.6 provides a summary of the chapter.

## 4.2 Data Collection

The data collection can be divided into two phases; initial and later approach which proved necessary as the system design changed in two crucial aspects:

- The prototype design suggested that the Exif information for each photograph would be sufficient in accurately describing the low-level characteristics of photographs.
- The prototype design was developed as a PHP-based image search using

MySQL DB. The complete system required that the images be processed to extract their dominant colors' color space and integrate SARA into the system and change the technologies used abandoning PHP and MySQL.

In this section the first two processes carried out as described in 4.2.1 and 4.2.2 were viable for both the prototype and the final implementation, but the later processes were only required for the complete system, hence were realized/performed only after the implementation of the prototype.

### 4.2.1 Photograph Repository

This research chose to utilize the Flickr photograph collection in order to build its own local cache. Flickr having a vast amount of photographs and an extensive open API led to this decision. In this research we aimed to follow an experimental approach where collective photosets would be explored in order to achieve a wide variety of non user-centric photograph collections, i.e. sets of arbitrary photographs with no distinct styles. Therefore, we needed to cache photographs from a large number of users. The ideal way to realize this requirement was to "ask the Flickr Pandas for a list of public photos". The Flickr API has two methods;

```
flickr.panda.getList
flickr.panda.getPhotos
```

The first call indicates which of the Flickr Pandas are available to request photos from. *Ling Ling* and *Hsing Hsing* return photos they are currently interested in, and both have different tastes in photos depending on their mood. *Wang Wang* returns photos that have recently been geotagged. The terms here "interested in" and "taste" aren't clearly stated by Flickr: "...because of the complexity and amount of stuff going on, no-one really knows exactly what Ling Ling and Hsing Hsing will find interesting,...".

The second call returns a list of photos that the Flickr Pandas are currently interested in, in the following format:

```
<photos interval=”60000” lastupdate=”1252358876” total=”120”
    panda=”ling␣ling”>
        <photo title=”P072909PS−0440” id=”3817425917”
            secret=”41991e776e” server=”2549” farm=”3”
            owner=”35591378@N03” ownername=”The␣Official␣White␣House␣
            Photostream”/>
        <photo title=”Amazing␣Sky” id=”3898206970” secret=”72e127e8f9”
            server=”2622” farm=”3” owner=”86665756@N00”
            ownername=”Sarah␣Ackerman”/>
        <!−− etc −−>
</photos>
```


Listing 4.1: Example Panda Call Response

It is possible for a Panda to sort through around 150, 000 photos per day. This approach enabled us to cache versatile photographs in a short term. As the Pandas only sort through public photographs, the developed tool abides by the Flickr requirements and guidelines to respect any user-determined privacy setting. It is the Flickr account holder's responsibility to restrict access to their data, if necessary. We used the Python Flickr API interface [Stüvel, 2009] in order to access the Flickr API.

**Python** is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding; make it very feasible for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together [Python, 2009]. Python is considered to be very fast and easily manipulated when parsing XML files and as the Python Flickr API interface is actively developed keeping up-to-date with the Flickr API versions, the decision to use Python and the related interface was made.

Over a period of a week we cached more than 12000 photographs from Flickr in small, medium, and large sizes.

### 4.2.2 Metadata Repository

The Flickr API provides the following methods which we made use of in order to retrieve the metadata for the cached photographs; `flickr.photos.getInfo` and `flickr.photos.getExif`.

The first call returns information about a photo in the below format:

```
<photo id="3898206970" secret="72e127e8f9" server="2622" farm="3"
    dateuploaded="1252360468" isfavorite="0" license="4" rotation="0"
    originalsecret="ddb92989b3" originalformat="jpg" media="photo">
  <owner nsid="86665756@N00" username="Sarah Ackerman" realname="Sarah 
      Ackerman" location="New York, USA"/>
  <title>Amazing Sky</title>
  <description>Trump International. Sunny Isles Beach. Miami, Florida.
      9.3.09.</description>
  <visibility ispublic="1" isfriend="0" isfamily="0"/>
  <dates posted="1252360468" taken="2009-09-03 03:04:44"
      takengranularity="0" lastupdate="1252366240"/>
  <editability cancomment="0" canaddmeta="0"/>
  <usage candownload="1" canblog="0" canprint="0"/>
  <comments>1</comments>
  <notes>...</notes>
  <tags>
    <tag id="134766-3898206970-735517" author="86665756@N00"
        raw="Sunny Isles" machine_tag="0">sunnyisles</tag>
    <tag id="134766-3898206970-4861" author="86665756@N00" raw="Miami"
        machine_tag="0">miami</tag>
  </tags>
  <urls>
    <url type="photopage">
      http://www.flickr.com/photos/sackerman519/3898206970/
    </url>
  </urls>
</photo>
```

Listing 4.2: Example getInfo Call Response

Analyzing the information provided we decided on the data that was necessary to be kept and formed the model 4.3 to store this information.


```xml
<?xml version="1.0" encoding="UTF-8"?>
<photos>
  <photo>
    <id>3898206970</id>
    <owner nsid="86665756@N00">Sarah Ackerman</owner>
    <title>Amazing Sky</title>
    <description>Trump International. Sunny Isles Beach. Miami,
        Florida. 9.3.09.</description>
    <dates posted="1252360468" taken="2009-09-03 03:04:44"/>
    <notes>...</notes>
    <tags>
      <tag id="134766-3898206970-735517" author="86665756@N00"
          raw="Sunny Isles" machine_tag="0">sunnyisles</tag>
      <tag id="134766-3898206970-4861" author="86665756@N00"
          raw="Miami" machine_tag="0">miami</tag>
    </tags>
    <urls>
      <url type="photopage">
        http://www.flickr.com/photos/sackerman519/3898206970/
      </url>
    </urls>
  </photo>
</photos>
```


Listing 4.3: Example InformationDB XML

The second call's response varied among the photograph collection based on the camera, image type, user alterations, etc. As this data would enable us to create a uniform language applicable to all photographs it was necessary to create a homogenous model. Parsing all the responses we formed the model 4.4 to store this information. Among our collection a small number of photographs did not contain this information in their headers; hence we disposed of those not fitting the model and ended with a final number of 12000.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<photos>
  <photo>
    <id>3898206970</id>
    <flash>No</flash>
    <make>NIKON CORPORATION</make>
    <model>NIKON D300</model>
    <exposureTime>1/800</exposureTime>
    <apertureValue>4.0</apertureValue>
    <whiteBalance>Auto</whiteBalance>
    <contrast>High</contrast>
    <saturation>Normal</saturation>
    <sharpness>Normal</sharpness>
    <iso>400</iso>
    <temperature>2950</temperature>
    <tint>+37</tint>
  </photo>
</photos>
```

Listing 4.4: Example ExifDB XML

Once the necessary data for each photograph was parsed and stored, we created the tags repository based on unique tags within the collection. Calling the `flickr.tags.getRelated` function, we received a list of tags related to the given tag, based on clustered usage analysis. In Flickr tags can be respresented within different clusters. For example when the `flickr.tags.getClusters` is called with the parameter "tiger" there are four different clusters returned:

**cluster 1** zoo, cat, animal, animals, wildlife, nature, bigcat, stripes, white, wild, ...

**cluster 2** orange, flower, lily, black

**cluster 3** mac, apple, osx

**cluster 4** swallowtail, butterfly

Analyzing this cluster use, Flickr provides the related tags. 4.5 shows how each tag and its related tags were stored.

```
<?xml version="1.0" encoding="UTF-8"?>
<tags source="train">
  <tag>station</tag>
  <tag>railroad</tag>
  <tag>railway</tag>
  <tag>rail</tag>
  <tag>tracks</tag>
  <tag>bw</tag>
  <!-- etc -->
</tags>
```

Listing 4.5: Example TagRepository DB

Having gathered all the required information, we dumped the data into a MySQL database and implemented the prototype. In order to create the expert vocabulary we started experimenting with the metadata, seeking a set of rules that would enable us to obtain an aesthetic oriented vocabulary. The metadata alone was not feasible when aesthetic perception was considered. Only objective concepts such as when and where the photograph was taken and whether it was an indoor or outdoor photograph could be derived. Therefore, we chose another approach in order to access the low-level features of a photograph. The language that color theory provides was taken into consideration at this stage. As it is both highly perceptive and generalized, it would form a strong base for an aesthetic semantic space.

### 4.2.3 Processed Content Repository

A digital photograph's content recorded by the CCD in pixel values was analyzed in order to get the dominant tones and colors. To find the dominant colors building a histogram would suffice, but also the dominant tones are of importance in this research as the most frequently used colors may not be the dominant tones. Also we needed to process the hue with a finer quantization as the human perception is more sensitive to hue than saturation and lightness. The process that the photographs were put through to extract their color space will be described by a digital photo-

**Figure 4.1**: The original digital photograph

graph[1] from the local collection shown in figure 4.1. First we divided the image into 5x5 blocks. This was necessary in order to preserve the region of the color while providing a fast analysis; each block was processed on its own. For each block we created new images which is shown in figure 4.2.

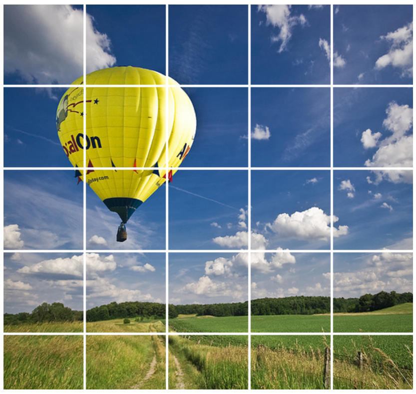


**Figure 4.2**: Divided into regions

Focusing on an individual region's pixel data, we extracted the red, green and blue values for each pixel. Then within each block these values were rounded to their contextual RGB values to avoid almost duplicate colors. Once each block was processed they were joined together to form a flattened copy of the original photograph. From this copy, represented by the HSL color (converted from the RGB model), the dominant ten colors were extracted. As figure 4.3 shows, the saturation and lightness of the colors were in some degree rounded, but the hue was preserved.

[1]Dreams of Summer by Bob Voors, `http://www.flickr.com/photos/voors/3187724878`

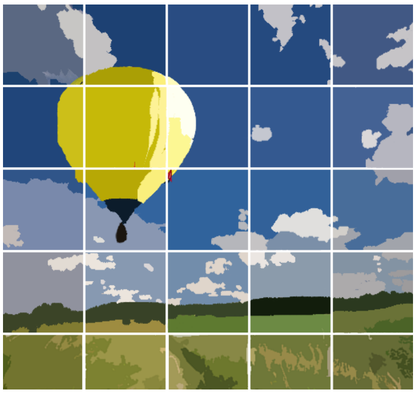


**Figure 4.3**: Colors after processing

```xml
<?xml version="1.0" encoding="UTF-8"?>
<colors>
  <color>
    <hue>220</hue>
    <saturation>60</saturation>
    <lightness>32</lightness>
    <count>1074</count>
  </color>
  <color>
    <hue>210</hue>
    <saturation>33</saturation>
    <lightness>39</lightness>
    <count>768</count>
  </color>
  <color>
    <hue>210</hue>
    <saturation>50</saturation>
    <lightness>26</lightness>
    <count>756</count>
  </color>
<!-- etc -->
</colors>
```

Listing 4.6: Example ContentRepository DB

The extracted data forms the model 4.6 which is stored in the content repository.

here count refers to the amount of the HSL measures' in a photograph based on pixels.

### 4.2.4 Domain Model

Having gathered all the data, we merged them in order to produce the final XML files that contain the photograph information, Exif and content data.

As mentioned in chapter 3 the system required the services of SARA. The first step to be taken in order to work with SARA is to construct a domain model. In compliance with SARA's XML Schema we created the domain model which, in essence, contains the metadata that the domain expert can use to manipulate and create semantic attributes. A simplified domain model which consists of only an example from each three sets (photograph information, Exif and content data) can be found in Appendix B.

## 4.3 Prototype

A PHP-based image search that supported a MySQL database was initially implemented to realize the requirements of the system and experiment with possible semantic attributes and their underlying features. This prototype did not focus on the discovery space, but rather enabled us to come up with the final domain model and the semantic attributes that would be encoded into the system based on these low-level data. Figure 4.4 shows a screen shot of the prototype. Here the collection represents the "Strong", "Content", "Deep" and "'Cool' photographs.

### 4.3.1 The Semantic Attributes

The expert vocabulary consists of nine sets of semantic attributes. Each set represents a key semantic attribute. The semantic attributes were created and encoded into the system via SARA's authoring tool. Here we will present the translation of low-level features into these key semantic attributes based on the examples "Cool" and "Strong".

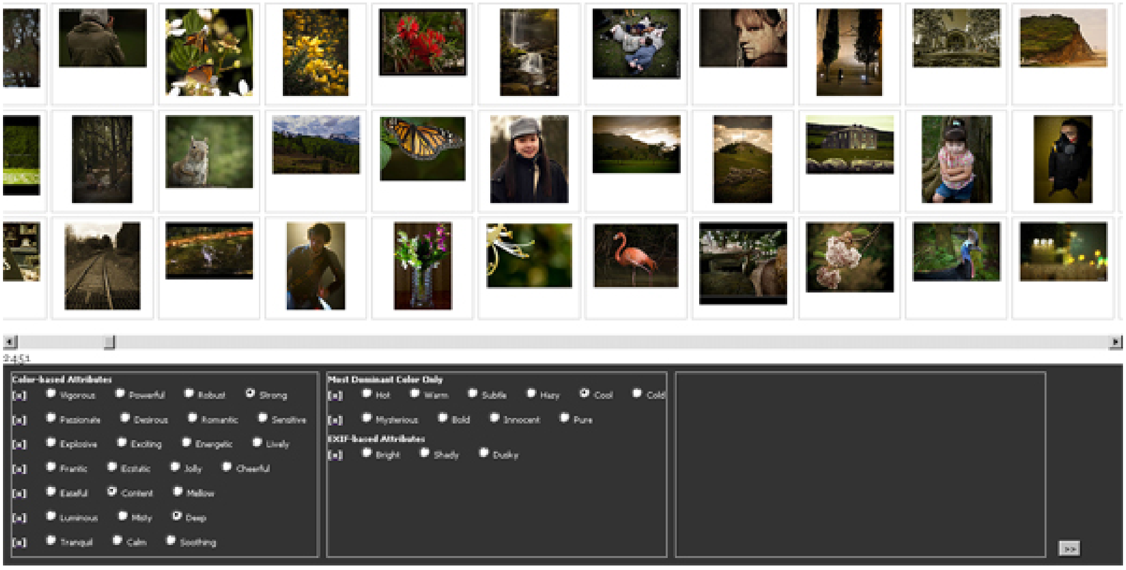



**Figure 4.4**: PHP-based image search

### Cool

Cool is a parameter of the semantic attribute "Temperature". When Temperature was defined, the hue of the color was taken into consideration. Within the color theory, colors such as blue and green are cool colors where red and orange are warm. Hue is represented as an angle of the color circle. So if we divide it into twelve equal interval; on each 30° angle we will have the colors; red, red-yellow(orange), yellow, yellow-green, green, green-cyan, cyan, cyan-blue, blue, blue-magenta, magenta, and magenta-red. We regarded red, yellow, green, cyan, blue, and magenta as the key colors having the intermediate colors in between. The classification for Temperature was then constructed as follows:

$$\{(0 \leq H < 75 \wedge 15 \leq L \leq 90) \vee (H \geq 300 \wedge 65 \leq L \leq 90)\} \wedge S \geq 25 \Rightarrow Warm$$
$$(75 \leq H < 120) \wedge (15 \leq L \leq 90) \wedge S \geq 25 \Rightarrow Subtle$$
$$(120 \leq H < 210) \wedge (15 \leq L \leq 90) \wedge S \geq 25 \Rightarrow Cool$$
$$(210 \leq H < 300) \wedge (15 \leq L \leq 90) \wedge S \geq 25 \Rightarrow Cold$$

where H,S and L respresent hue, saturation and lightness respectively.

If a photograph's color space satisfies the third equation, it will be considered as a Cool photograph and the degree will depend on the Count.

**Strong**

In this vocabulary strong is a parameter of the semantic attribute "Power". This semantic attribute is based on the idea that warm colors can evoke enthusiasm or anger, i.e. warm colors express emotions from light optimism to strong violence. Color psychology also associates the color red with power. Combining these two ideas we considered colors with hues in the interval [0,25] to be associated with power without allowing the "optimism" or "joy" of color yellow, as color psychology denotes them to be, into the picture. The level of power was then tweaked with saturation and lightness of the color. A really vibrant and saturated color would be Vigorous where a less impactful, dim color would be Strong. Hence vigorous and strong were defined as follows:

$$\{H \leq 25 \wedge (75 < S \leq 100) \wedge (40 \leq L \leq 60)\} \Rightarrow Vigorous$$

$$H \leq 25 \wedge \{(50 \leq S \leq 75 \wedge 30 \leq L \leq 60) \vee (75 < S \leq 100 \wedge 30 < L < 40)\} \Rightarrow Strong$$

## 4.4 GUI Prototypes

Having specified an expert vocabulary we then focused on the front-end. Looking at the prototype and how the content spread, the general drawbacks were acknowledged:

- Crowded tag space
- Flat attribute space
- List-based photo collection presentation

All in all the front-end had limits in user interaction and had no aesthetic appeal. In order to address these issues each were tackled separately.

### 4.4.1 TagBall

As mentioned in chapter 3, the number of crowd sourced tags exceeded a number that could be clearly displayed with classic list-based presentations referred to as simple tag clouds. Therefore we chose to implement a tag space which would allow large numbers of tags to be displayed while not cluttering the UI. The tag space also had to be distributed on a geometry that could easily be manipulated and controlled by the end-user. Having the tags cover the faces of a sphere which could be resized and dragged addressed these necessities. The implemented TagBall within the front-end can be resized with the mousewheel and dragged by simply clicking and pulling it anywhere on the interface plane. The TagBall's size has no limits and can always be hidden by clicking on its icon button.

### 4.4.2 AttBar

The expert vocabulary's integration into the front-end is presented with the interface component called the AttBar. Multiple sets of semantic attributes are placed on this item where each set displays a semantic expression and its extended terms. The user can select one parameter from each bar to run a query. The end-user can decide if the AttBar will be visible or not by clicking on its icon button.

### 4.4.3 Discovery Space

Three prototypes of the Discovery Space were designed in order to address the requirements outlined in chapter 3. This section presents each of them along with the reasons why they weren't considered for the final version of the Discovery Space. The PHP-based prototype had a list-presentation in which photographs had to be browsed with a window-slider and its drawbacks have already been discussed in chapter 3, hence wasn't a part of the considered GUI prototypes.

**A conical spiral presentation of the photographs**, as shown in figure 4.5, was considered. The idea behind this design was to represent the relevancy of the photographs to the query by having the most relevant photographs on the outer scope and as the user progressed through the "tunnel" of photographs, they would become less and less relevant. The drawback of this design was that it did not portray the big picture; the users would only be able to see all the photographs as long as they were on the most outer scope and even then the less relevant photographs wouldn't be clear due to their size and the overlapping of the photographs. Hence this design was not considered for the final system.

**Figure 4.5**: A conical spiral presentation of the photographs

**A circular rotation of the photographs** was considered next, where the photographs would cover the inner surface of a cylinder and the user would browse through the collection by turning the cylinder. The cylinder could also be resized, i.e. the radius could be changed by the user depending on the delta of the mousewheel event the system received. Figure 4.6 shows the considered approach. The difficulty this design imposed was the dependency on the number of the result set. This number had to be considered when drawing the cylinder. Also as with the first approach, users wouldn't be able to see the whole sett without zooming in and out, which caused unnecessary interactions.

**Figure 4.6**: A cylindrical presentation of the photographs

**The interaction with individual photographs** was considered separately.

As shown in figure 4.7, the end-user was allowed to bring a photograph into focus by pulling it towards and could simply remove a photograph by pushing it out of the surface. In each case the photograph's size increased and decreased (ultimately the photograph disappeared from view) relatively. Although this approach more or less shaped the final design, later on in the implementation the removal of photographs was abandoned, hence this approach became obsolete.

**Figure 4.7**: The interaction with individual photographs

## 4.5 Final System

### 4.5.1 Technologies Used

This section outlines only the technologies that were specifically chosen to fulfill the requirements of the system. For example later on in the project, the system was migrated to the Flex framework, but it was a trivial requirement.

#### Adobe Flash - AS3

Adobe Flash is a multimedia platform which enables the delivery of rich and interactive content by adding animation and interactivity to applications. In this implementation ActionScript 3.0 (AS3) a scripting language of Adobe Flash was used to create the user interface. AS3 is an object oriented programming language hence allows more control and code reusability when building sophisticated Flash applications.

#### Papervision3D

Adobe Flash has limited 3D capabilities. The latest version; Adobe Flash CS4 Professional comes with only two 3D tools, the 3D Rotation tool (for rotating symbols in a 3D space) and the 3D Translation tool (for moving symbols in a 3D space). In order to achieve real life like 3D effects this implementation used the Papervision3D (PV3D) engine [PV3D, 2009]. PV3D is an open source high performance 3D engine for Flash. With PV3D we created advanced three-dimensional objects to offer a high visual experience and interaction to the end-users. Through this interactive environment the content exploring process was enhanced and an engaging highly visual user interface was developed. Based on the available comparisons of PV3D and other 3D rendering engines (such as Sandy and Away3D), PV3D was found to be suitable for this implementation as it offered better performance, more extensive classes and large online documentation.

### eXistDB

eXist-db is an open source database management system built using XML technology. It stores XML data according to the XML data model and offers index-based XQuery processing [eXistDB, 2009]. In this implementation the processed content along with the retrieved photograph information and Exif data are stored as XML files in an eXist database.

### AMFPHP

AMFPHP is an open-source PHP implementation of the Action Message Format (AMF). AMF allows for binary serialization of Action Script (AS2, AS3) native types and objects to be sent to server side services [AMFPHP, 2009]. AMFPHP allows thin client applications that are built in languages such as Flash and Flex to communicate directly with remote PHP class objects. Using AMFPHP, intensive methods that were considered somewhat tedious and considerably lengthy were offloaded to the local PHP server. Via AMFPHP, the database connection was also left to the PHP remote object. AMFPHP was an efficient solution for this implementation as it is suitable for simple data requests, e.g. a datagrid/table population from a database. Also speed was an important necessity in this system in order to maintain a fluent user experience. AMFPHP serializes the communication into a binary format making it a fast client server communication protocol.

### PheXist

This implementation used PheXist which is a set of classes, implemented in both PHP and Perl, that allow querying the eXist database. Connection with the eXist database is done through a SOAP interface, using a WSDL definition [PheXist, 2009].

### 4.5.2 Architecture

Figure 4.8 shows the overall architecture of the application. The SARA process model given in the shaded area is a highly summarized model of the actual flow and does not portray the definite framework. Further in depth architecture can be found in [Hampson, 2009].

The system flow was described in chapter 3. In this implementation the client application talks to the apache server via the AMFPHP object which queries the eXist database via PheXist. The calls to SARA are made through the remote SARA object.

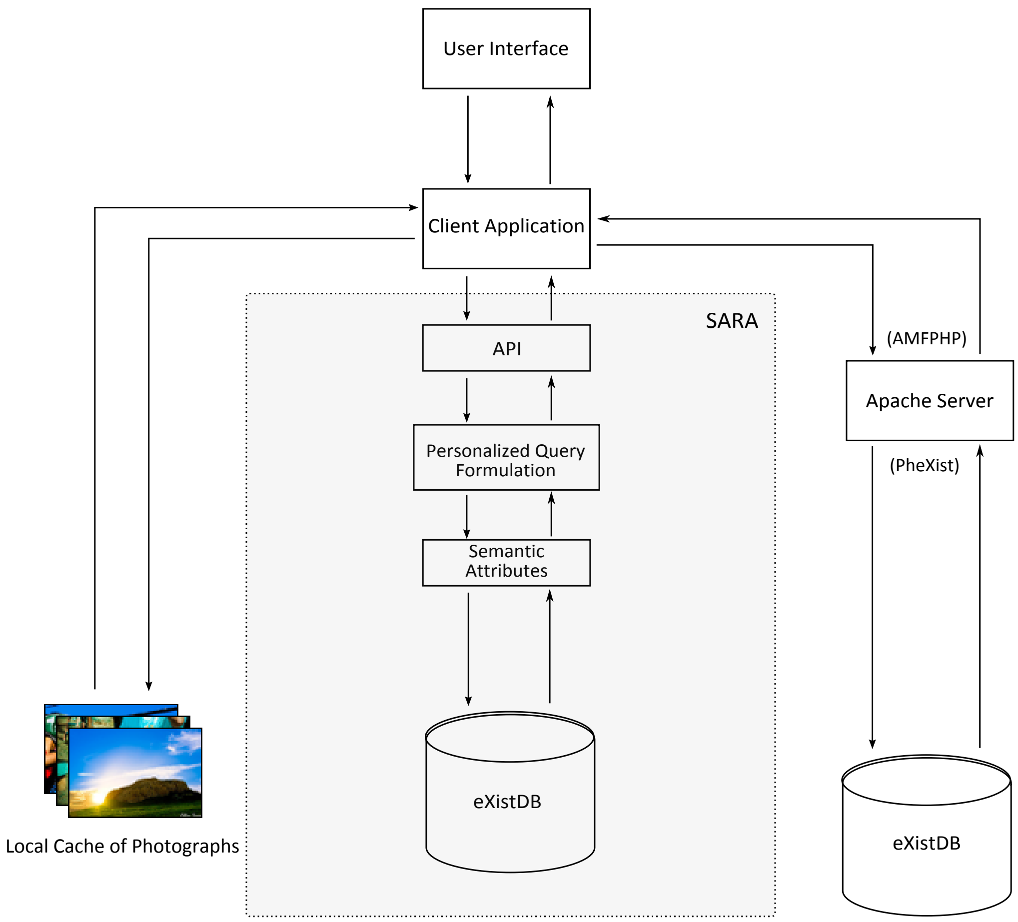



**Figure 4.8**: Final System Architecture

### 4.5.3 A demonstration of $X^2Photo$

Figure 4.9 shows the front-end of the system in which case the user has carried out a query choosing the semantic attributes “Lively”, “Luminous” and “Cool” from the AttBar. The photographs are displayed in the Discovery Space while the TagBall is populated with the popular tags of the resulting collection.


**Figure 4.9**: Exploration

When the user clicks on a photograph the interface zooms in to the photograph as shown in figure 4.10.

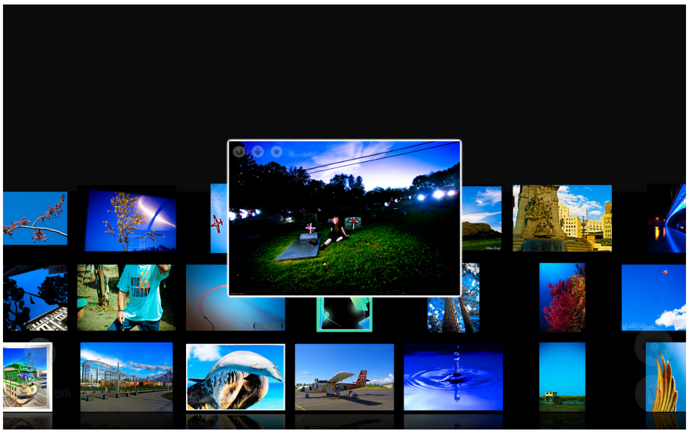


**Figure 4.10**: Zoomed in View

Here the user can click the flip button on the top left corner of the photograph to see its details, presented in figure 4.11.

If the user clicks the down arrow the photograph will be brought into focus, displayed in figure 4.12. Once the photograph is in focus; the AttBar displays its

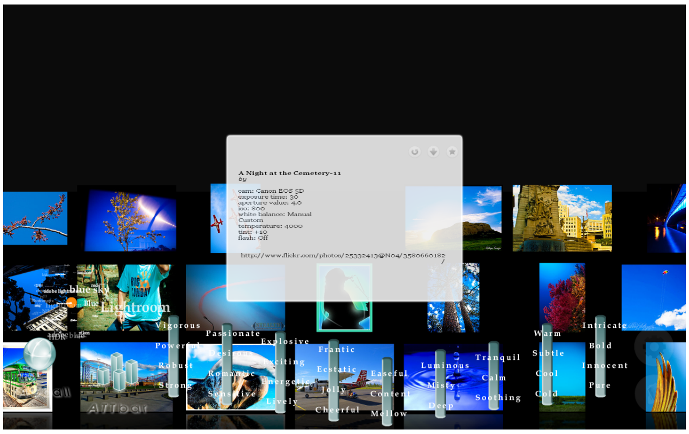



**Figure 4.11**: Viewing Details

semantic attributes accordingly and the TagBall now consists of its tags only. In this case the particular photograph has no tags and is considered to be a "Content", "Luminous", "Tranquil", "Subtle" and "Intricate" photograph by the system.

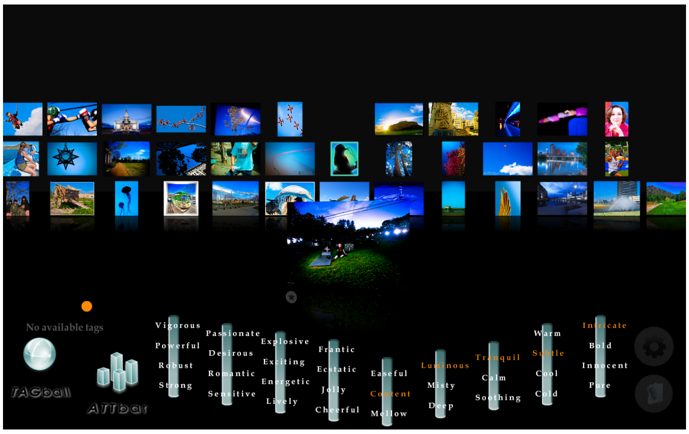



**Figure 4.12**: In Focus

The user can send the photograph to the Favorites area, shown in figure 4.13 by clicking the star button either when the photograph is zoomed into or brought into focus. The Favorites area will be accumulating the photographs as the user sends any into this area. In this area the user can also bring a photograph into focus to see its related tags and attributes as shown in 4.13.

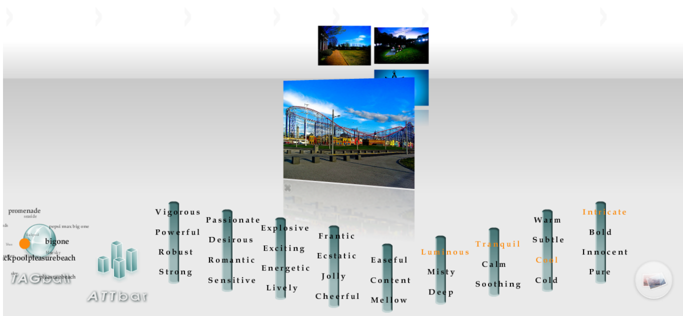



**Figure 4.13**: Favorites Area

### 4.5.4 Analysis

$X^2$*Photo* aims to be a tool that enables users to retrieve photographs from large collections using not just objective tags but also high-level semantic attributes that are created by experts who base their subjective expressions on low-level image contents, the color space in this case. In order to achieve this goal and offer good performance the system had certain requirements to fulfill.

To make use of tags, the system offers the TagBall which enable end-users to view the tags associated with photograph collections obtained from query results as well as each individual photograph in order to define the labels for consecutive refined searches.

The color theory is the basis of the offered AttBar, which displays the semantic attributes created by the expert via the authoring tool that SARA provides. This asset enables the end-user to retrieve photographs by more intuitive expressions without limiting them to the content of the photograph.

The implemented tool, as shown in the figures above, has a highly visual and attractive user interface which enables end-users to interact with the system through several provided access points. Each element and their combination offer the user a rich and highly interactive exploration of photographs.

The system introduces the concept of refining a search focusing on a particular photograph. End-users can carry out consecutive searches based on a specific photograph. This aspect enables them to express their target more efficiently and also gain an idea of the expert's perspective.

## 4.6 Summary

This chapter detailed the implementation process following a chronological order. The gathering procedure of a 12000 photograph collection and their accompanying information were described, and how the collected data was parsed to create the static sources. Early prototypes which led to the development of the final system were presented. After demonstrating a flow of the system, we showed how the system met the requirements brought up in the design chapter.

# Chapter 5

# Evaluation

## 5.1 Introduction

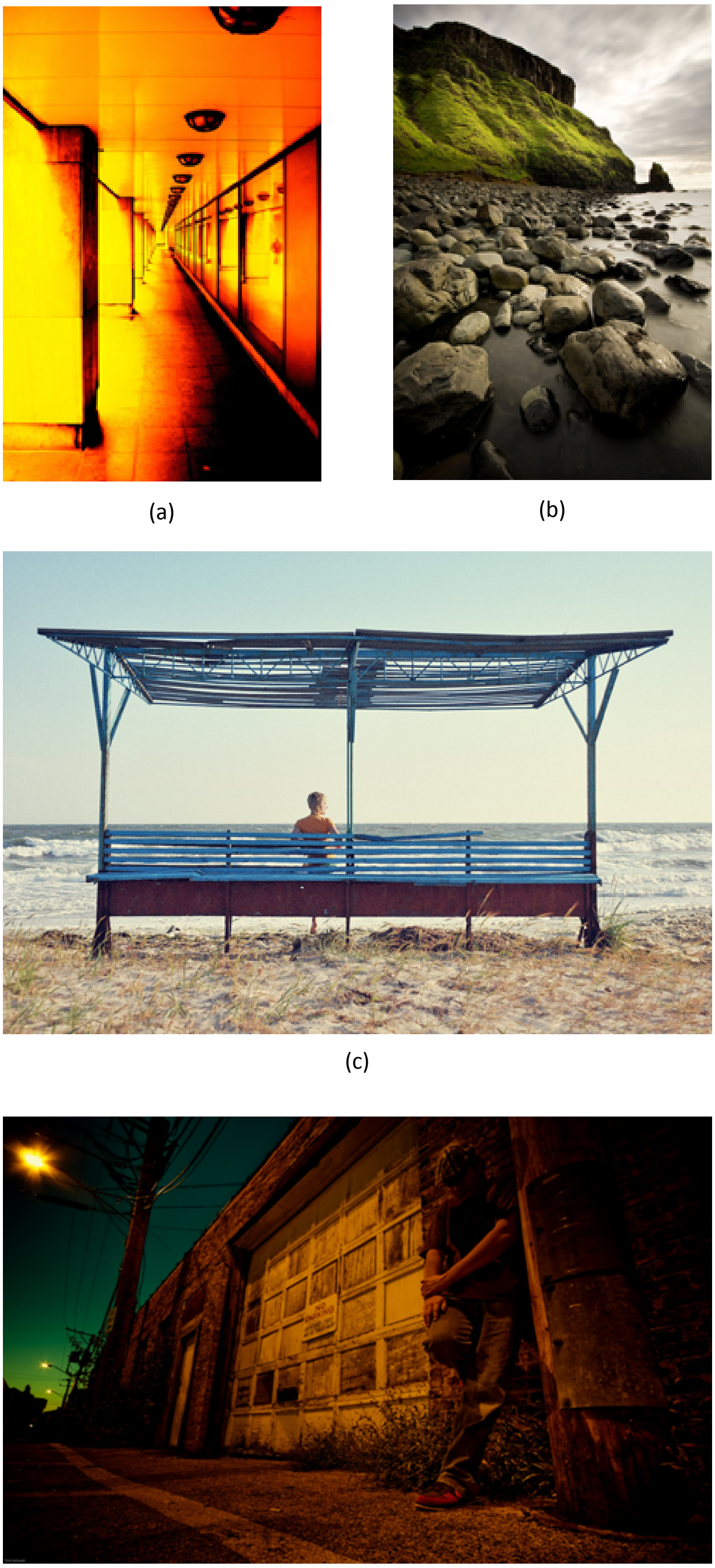
This chapter presents the results of the usability tests carried out in order to evaluate $X^2$*Photo*. Section 5.2 discusses the small-scale user study conducted, outlining the evaluation setup and the users involved. In section 5.3 the test process of the user group is presented along with some remarks. Section 5.4 outlines the results of a survey conducted to obtain user assessment and feedback. The accompanying questionnaire can be found in Appendix A. Section 5.5 analyzes the user evaluations considering the user tests and the feedback given, and section 5.6 gives a summary of the chapter.

## 5.2 User Study

### 5.2.1 Evaluation Setup

The purpose of this evaluation was to test the usability, functionality and the overall appeal of the implemented tool as well as to investigate the potential benefits of injecting subjective expert knowledge based on the manipulation of non-textual low-level data when compared to conventional methods, such as image retrieval via tags only. With this aim in mind we pursued the following approach:

Four photographs shown in figure 5.1, not present in the 12000 photograph col-



(a) (b) (c) (d)

**Figure 5.1**: Photographs shown to users

lection were selected based on different criteria. All the photographs were expressed with different aesthetic semantics via $X^2Photo$ and the range of the vocabulary covered many available attributes within the system. Also our intention was to produce photographs which would allow different perceptions. For example while 5.1(a) was considered a somewhat abstract photograph, 5.1(b) and 5.1(c) were clear in pictorial content. 5.1(d) had a story-like content which could be interpreted differently by individuals.

Each user was first shown all four photographs which they were asked to describe in their own words. Then they were given an overview of the tool with similar screenshots presented in the demonstration in chapter 4.

Once the screenshots were shown they were exposed to the tool and were asked to do the following tasks:

- For photograph 1, find similar photographs via $X^2Photo$
- For each photograph found, add it to the Favorites
- Repeat this task for all four photographs.
- Once complete, go to Flickr and for photograph 1; again try to find similar ones either with the words originally used to describe the photographs or with different ones.
- Repeat this task for all four photographs.

The photographs that the users would find via Flickr were to be collected by taking screen shots of the Web browser. After finishing these tasks they were given a survey to fill out, which is discussed in 5.4, to complete the user-test.

### 5.2.2 User Group

Nine users were selected for this study where seven were from the Networks and Distributed Systems M.Sc. course at Trinity College Dublin. Four users indicated that they were interested in photography and considered themselves amateur photographers. Only two users didn't have any online photograph sharing site accounts, one stating that he does but is not using them as he "couldn't find value in them", and the rest of the users preferred either Picasa (4/7) or Flickr (3/7) to share their photographs. Tagging was not a shared habit among the users, while only one used the method to mostly associate photographs with labels concerning the content of the photograph or the location it was taken. Three users indicated that they had personal blogs through which they display their photographs and tend to caption them with people and location or small notes. Users collectively stated that they use Google Images, while two people also preferred Bing and Yahoo! Images primarily to search for images on the Web. One user stated that he used Cooliris as an interface to access Google Images. Two users indicated use of Flickr and one other iStockPhoto. Two users had been exposed to the tool before, and they will be mentioned accordingly when necessary.

## 5.3 User Tests

The way how the users described the four photographs had some interesting aspects. Users who were interested in photography tended to use more technical terms. For example, they used words such as "over-exposed" for the first photograph and observed how the third photograph might have been shot talking about the angle of the camera and questioned whether it was altered in an image editing program to obtain the deep contrast. They didn't consider the content as much when compared to the other users. Some users preferred to describe the photographs with more personal expressions such as "lonely" and "tempting" when referring to the second photograph. The fourth photograph, as expected, was interpreted differently by almost all the users. While some tried to figure out what the man in the picture might be doing, some chose to describe him but again had different impressions such as "gritty", "relaxed" or "run-down". Almost all the users chose expressions like "warm", "cold", "airy", "gloomy" and "energetic" which was interesting to see how some of them coincided with the actual attributes determined by $X^2Photo$. They usually chose to first define the photographs with such expressions and then focused on the actual content. Two users were more objective in their descriptions and chose to name the elements they saw in the photographs with words like "corridor", "bench", "rocks", "back alley", etc. Majority of the users combined their perceptions with the content: "...a cool calm pic but alive. There's a woman sitting on a bench... feels breezy but soft... waves look relaxing".


**Figure 5.2**: Similar photographs found using only the AttBar by a disagreeing user

Just like their preferences in describing the photographs, their approach to find similar photographs were particularly different. Out of the nine users that were interviewed, four never actually enabled the TagBall. Coincidentally their descriptions of the photographs were heavily consisted of expressions like "moody", "dark", "calm", etc. They directly chose similar words present within the AttBar and carried out their searches. With the first results they received two users were surprised to see how the tool interpreted their descriptions. They did not agree with the expert and started experimenting with the AttBar rather than continuing with their searches. After observing some consecutive result sets and bringing some photographs into focus, they grasped the association and modified their searches accordingly. Figure 5.2 and 5.3 shows the collective photographs these two users found.

The other two users only performed 2-3 consecutive searches and by refining their search found the similar photographs shown in figure 5.4.

Observing the photographs it is interesting to see what the users based their similarity criteria on. While some photographs have a similar feel to them regarding the concept or the context, some are similar in content as well.


**Figure 5.3**: Similar photographs found using only the AttBar by a disagreeing user

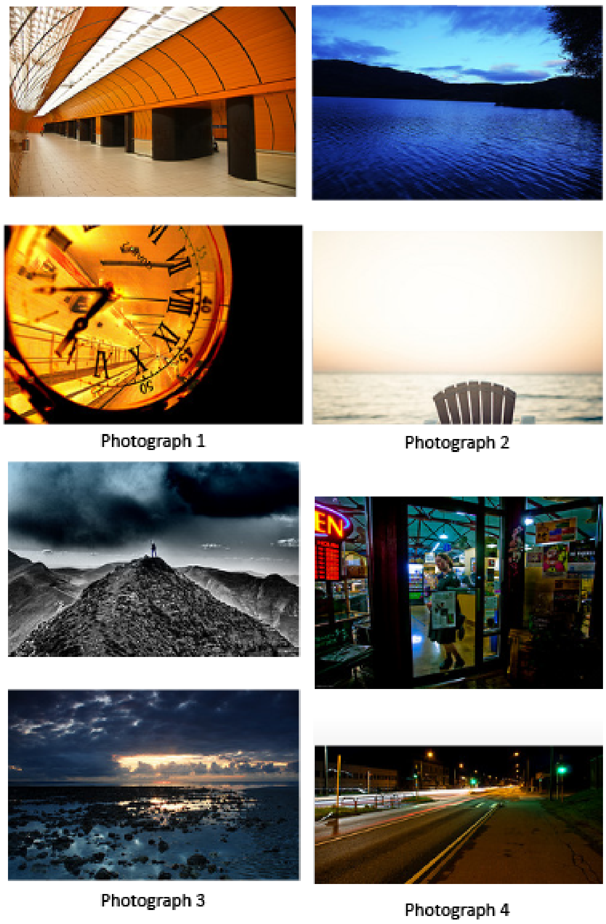



**Figure 5.4**: Similar photographs found using only the AttBar by agreeing users

Within the remaining 5 users only one user started off with approaching the TagBall first. For example to find similar images like the second photograph the user chose “beach”, “girl”, “sky” and “nature”. The result set he received was rather an extensive one and the reason for this is much self explanatory when we observe the returned tags related to “sky” from Flickr. Especially the two tags “nikon” and “canon” are enough to retrieve half of the local cached 12000 photograph collection as they are the most popular camera models among Flickr users [CameraFinder, 2009].


**Figure 5.5**: Similar photographs found using only the TagBall

It was observed that the system needed performance adjustments in order to support vast number of photographs. The user did manage to find similar photographs, but stated that the system wasn't empowering as much he would like it to be and that the slowness and the retrieved photograph collections made the system frustrating at times.

When the task was complete the user was reminded of his original descriptions of a photograph where he used “misty” and “warm” for different photographs, and when asked why he didn’t adjust the AttBar using these semantic attributes, he stated that he did not use the AttBar or bring a photograph into focus as the procedure was rather confusing and that he was “lost at the beginning... didn’t know how to continue from the focus photo vocabulary sometimes unclear ... maybe too many points of entry at once...”. The photographs that the user found to be similar are shown in figure 5.5.


```
<tags source="sky">
  <tag>clouds</tag>
  <tag>blue</tag>
  <tag>sunset</tag>
  <tag>water</tag>
  <tag>sun</tag>
  <tag>trees</tag>
  <tag>landscape</tag>
  <tag>sea</tag>
  <tag>night</tag>
  <tag>sand</tag>
  <tag>beach</tag>
  <tag>tree</tag>
  <tag>green</tag>
  <tag>orange</tag>
  <tag>red</tag>
  <tag>nature</tag>
  <tag>moon</tag>
  <tag>summer</tag>
  <tag>storm</tag>
  <tag>grass</tag>
  <tag>reflection</tag>
  <tag>light</tag>
  <tag>cloud</tag>
  <tag>mountains</tag>
  <tag>evening</tag>
  <tag>flower</tag>
  <tag>canon</tag>
  <tag>sunrise</tag>
  <tag>black</tag>
  <tag>nikon</tag>
  <tag>ocean</tag>
  <tag>river</tag>
  <tag>winter</tag>
  <tag>boat</tag>
  <tag>lake</tag>
  <!-- etc -->
</tags>
```


Listing 5.1: Flickr response returning the tags related with sky

The remaining 4 users chose to use both the TagBall and the AttBar. It should be noted that two of these users had prior exposure to the system and were familiar with the expert’s view. They both had an easier and smoother experience and found plenty of quite similar photographs, some shown in figure 5.6, utilizing all the features of the system.

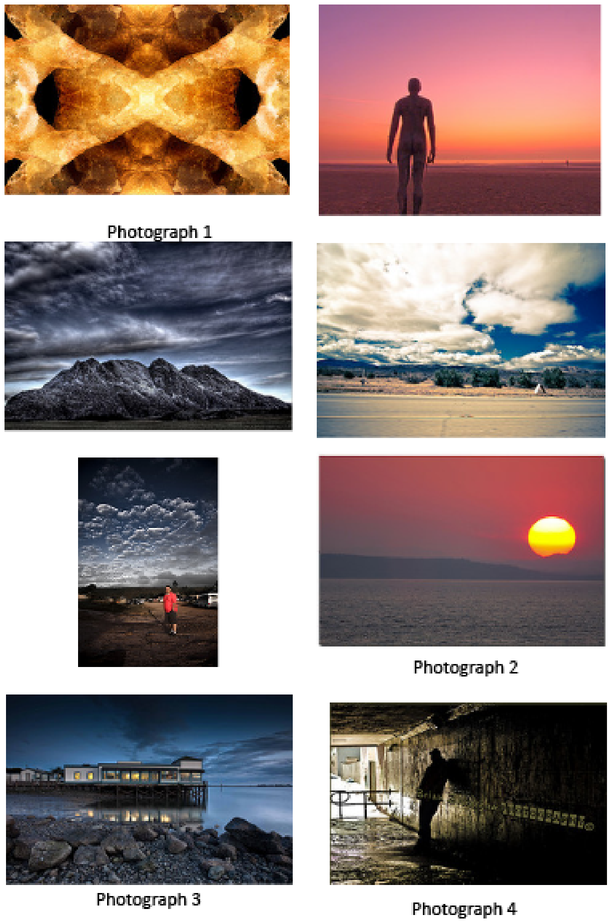



**Figure 5.6**: Similar photographs found using all features

When compared to the previous user, it might be possible to say that the system might require some time to ease into it if not clear at first. Within the other two users, one chose to make slight refinements and then “free-cruise” the collection rather than narrowing down the result set to find a wide set of photographs. When asked how many pages of images he would normally browse in image search engines he replied 2-3. As he went through a rather large amount of photographs, refining the set very rarely indicates that the Discovery Space within the tool supports free browsing as the procedure is not broken with clicking to refresh the view. The other user chose to search with the AttBar and occasionally the TagBall, choosing one or two specific tags.

Next when the users tried to find similar photographs in Flickr, their approach was again different at first. For example one user started with “fiery clinical harsh” to search for the first photograph, which were the expressions he had used when describing the photographs, while another directly abandoned his expressive vocabulary used originally based on his prior experience with Flickr and chose to use the search phrase “Scotland cliff coast”. Some users were very articulate in their searches and submitted phrases such as “city lights low angle journalistic lonely” to find similar photographs. During 3 user’s searches within Flickr a slight change in their vocabulary was seen. A user who had previously described the second photograph mainly based on content; “beach, person sitting on the bench, grayish” found a similar image within the implemented tool that the expert thought to be “romantic”, “soothing” and “innocent”. The user carried out his first search with “romantic sea scenery”.

Within the Flickr searches it was most interesting to see how the user’s search terms changed significantly. However the users started their search they met at the similar keywords, sometime with even the exact ones. For example the user that started with “fiery clinical harsh” carried out the consecutive searches as follows:

- fiery clinical harsh - found no results
- fiery - browsed 3 pages found no results
- fiery dead - the user pointed out that some photographs of Barack Obama, the US President were also included in the result set
- fiery dead hall - found no results
- fiery building - found the photograph shown in figure 5.7.

Two users ended up finding photographs they thought to be similar with the same terms “orange corridor” even though one started out with “warm burn hallway” and the other “energetic university bunkers”. Also for the second photograph the users almost always changed their terms to finally submit “bench beach”. The user who had started with “romantic sea scenery” changed his terms as follows:

- romantic sea scenery - found no results

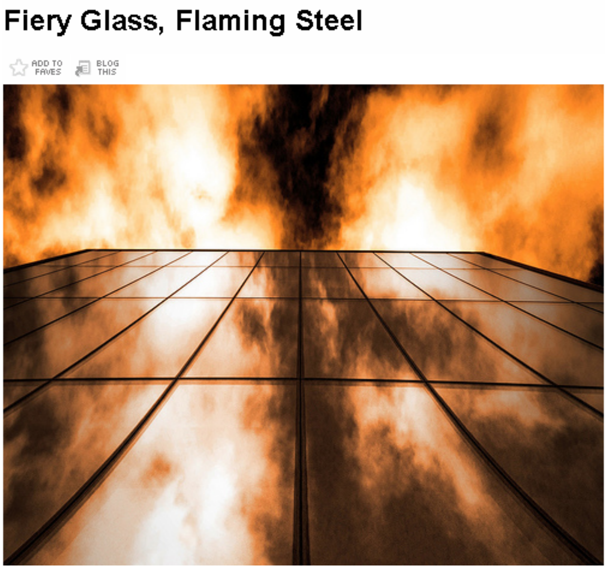



**Figure 5.7**: Similar photograph found in Flickr by a user

- sea scenery - found no results
- romantic beach - found no results
- relaxing beach - browsed 3 pages
- bench beach - found a similar photograph shown in figure 5.8.

The same term was used by 4 other users changing slightly; adding “sea”, “cold” or “sand”.

When looking at the photograph it is possible to see the same elements that are present in the shown photograph, but the similarities in photographic value and aesthetic qualities are questionable as pointed out by the user himself. Users familiar with Flickr also used the advanced search available and refined their queries, but again used content-based terms to carry out their searches. For example for the fourth photograph a tag oriented user chose the keywords “city night backstreet USA”. The terms chosen are not all related with the photograph, but the user assumed that he would find similar photographs with these keywords. Being exposed only to the photograph shown in figure 5.9 he searched for “backstreet USA”. The photographs he found similar are shown in figure 5.10. In the end all the users were able to find at least one similar image, which was not surprising at all considering


**Figure 5.8**: Similar photograph found in Flickr by a user

the vast amount of photographs Flickr has (as discussed in section 2.3.2). A screen shot was taken of the similar photograph(s). It was interesting to see how all the users had to resort to content-based terms, and in most cases the same ones (as with "orange corridor" or "beach bench"): This showed how such systems disregard the way each individual perceives a photograph to be, how they describe it, and ignore dissimilar appreciations.

## 5.4 User Survey

A user survey was conducted using an evaluation questionnaire when the given tasks were completed. The questionnaire intended to evaluate each feature's functionality as well as aesthetic qualities, and also the overall system quality regarding various aspects.

The general response to the usability and appeal of the Discovery Space was very positive, agreeing that the continuous flow enabled them to browse the photographs more thoroughly and that the interaction with the space was appealing; being able to go back and forth between the photographs by dragging the plane was considered


**Figure 5.9**: A photograph found in Flick with the tags "city, night, backstreet, USA"


**Figure 5.10**: Similar photograph found in Flickr by a user

as an engaging activity. One user did suggest that having scroll buttons on either side of the Discovery Space might be useful for less experienced/able users when browsing through the photographs.

The interaction with individual photographs within the Discovery Space; zooming in to view a photograph, flipping it to see the extra details and bringing it into focus, were evaluated individually. The zoom effect was found to be very good (5/9) or good (3/9), although one user who found it poor suggested that it could be more flexible: Rather than giving an option of one level of zoom, the users could zoom in and out with various levels, a flexibility similar to that of the TagBall. The extra details observed when the photograph is flipped was regarded as useful, especially by the users who considered themselves photograph enthusiasts: 3 users stated that being able to see the optical settings (such as the exposure time and aperture value) would give them an idea about the settings for certain shots they liked. 8 users were in favor of the functionality to bring a photograph into focus, and one person suggested combining the zoom and focus: When in zoomed view the TagBall and the AttBar could be populated accordingly without having to change the view a second time. The same user then also pointed out that it may not be possible to clearly see the related attributes but that could be overcome by enabling a multi-level zoom effect.

When asked how the users found the usability of the TagBall, the responses varied from very good (2/9), good (5/9) and to poor (2/9). Within the two users who regarded it to be very practical, one stated that being able to "go through all those popular tags" gave an idea about the collection and that it helped them to find the tag they were looking for as they "couldn't actually remember that word". One user also pointed out that it was good for random searches in which the purpose is to just browse through the collection with a general idea of the tags contained within. On the other hand the two users who found the TagBall to have a poor usability said that it was time consuming and that not being able to manually enter a keyword in an input box was restricting. Although they said that it looked good and melded in finely with the UI, they thought it was necessary to give the users an option to whether use the TagBall or an input box.

The AttBar was thought to be very good (8/9) and good (1/9), that the concept was comprehensible and the classification of the attributes were clear. The general idea was that it was an interesting asset within the whole UI.

Similar with the AttBar, the Favorites area received very good feedback. In fact, the visual appearance was considered to be better than the Exploration area. The switch between the two areas was especially appreciated. An interesting functionality one user suggested was the addition of an extra feature where the system would search for similar photographs according to the favorite photographs pool which the users accumulated in time or suggestions for similar images within the system when the user pulled a photograph into focus within the Favorites area.

Almost all the users considered the overall UI to be very good (8/9). Users often stated that it had a "slick and shiny feel to it" and found it to be intuitive. One user suggested that having mouseover tooltips for each button and also an intro screen might be more useful. The user thought that the interface was confusing at times since there were "too many entry points at once", referring to the TagBall, AttBar and in focus image.

As mentioned in the previous section the users varied in their preference of search methods. Their answers given to the functionality of search coincides with the approaches they took.

Two users within the 4 that didn't use tags while searching for similar photographs didn't comment on the functionality of the TagBall. Rest of the user opinions ranged from very useful (2/7), useful (2/7) to not useful (3/7). The 3 users who found the TagBall to be not useful had carried out searches using only tags and only aesthetic attributes. The user who preferred to use only tags tried to find ones that described the content of the photograph not focusing on its aesthetic value or on expressive qualities. For example regarding the photograph shown in 5.1(c); the user selected "beach", "girl" and "nature" from the TagBall. When presented with the results, the user found them to be too scattered and questioned some of the photographs' relevance with the tags that were chosen. The other two users had actually carried out successful searches with using the AttBar only and while one stated that "tags aren't my forte (in general)" the other user thought that the

AttBar caught more attention and made the idea of searching with tags a secondary approach.

The AttBar's use when carrying out a search was considered to be very useful (3/9), useful (5/9) or not useful (1/9). Two users, even though they disagreed with the expert, thought that after some random queries it was possible to adapt to the expert's view and carry out searches with the AttBar accordingly and fount it to be very useful or useful. The user who considered the AttBar to be poor in a search stating that the "attributes were frustrating" thought that the idea of being able to change the underlying rules of the attributes might encourage him to use the AttBar more. The user also stated that some of the attribute names weren't clear. The majority of the users who found the AttBar useful thought that enriching it with more attributes would be a good idea.

One of the largely accepted functionality within the system was the ability to refine a search with a focus image; 7 users thought it was very useful while the others found it useful (1/9) and not useful (1/9). The user who preferred to search for a similar photograph by only utilizing the TagBall found the refining idea to be poor based on the fact that some images contained very few tags that could enable further search. Rest of the users appreciated the refining process and used it for all four photographs they were asked to find.

Within the questionnaire the users were finally required to evaluate the system based on different criteria. Majority of the users (8/9) strongly agreed that the system was in fact attractive, and one user thought that it could be enhanced with small alterations such as making it more informative and clear for the end-user. All the users agreed that the system was a powerful tool for exploring photographs. When asked to assess the rapidity of the system, 3 users thought it was slow, stating that the photographs sometimes took some time to load and that the system might be jerky relative to the amount of the photographs displayed. Only one user thought that the system was neither empowering nor responsive. The user, contrary to the majority, also answered that the system was confusing and frustrating, and that it was difficult to navigate without guidance; suggested the necessity of more definers within the UI as the learning curve might be somewhat steep. Although he found the individual system assets to be clear within, he thought that it was possible for the-end user to be unsure about how to proceed during exploration utilizing all of them. When asked if the system was extensive, the users strongly agreed (3/9), agreed (5/9) or disagreed (1/9).

## 5.5 Analysis

### 5.5.1 Overall Evaluation Results

The user test and the succeeding survey suggest that when describing photographs, whether interested in photography or not, people tend to communicate "how" a photograph is and then "what" it portrays. This finding indicates a need for a more free vocabulary to be accessed in order to retrieve accurate and relevant photographs from any collection. Traditional tag-based systems restrict users to content-based terms, thus ignoring the artistic quality which is the actual factor that evokes appreciative emotions. Hence traditional tag-based systems reduce photographs to simple words on paper. As most people have become accustomed to this approach, in such an environment they ignore the initial values with which they would approach a photograph and are therefore relegated to search for the tagged simplification of a photograph rather than the actual photograph itself.

Based on photographs found by the users, using more natural expressions via the system, indicate that this approach can be used to grant users the freedom they seek when relating to a photograph. Injecting expert knowledge based on the manipulation of raw low-level data into a conventional system running on tag-based search only, allows users to freely express both the photograph and the picture it conveys. Even though a specific expert vocabulary may not be suitable or correct for each individual, users can adapt to the expert's view or better yet choose to subscribe altogether to a different expert expanding the semantic space.

### 5.5.2 System-specific Requirements

$X^2Photo$ received overall positive feedback; the users clearly understood the idea and also the overall concept. They were suggesting to subscribe to different experts, bring in Favorites based search, add more semantic attributes, which indicates that the users understood the aim of the tool and considered it to be a tool that could be extended further. When they were asked if they could see a real-life application of the tool they all agreed if further improvements were made. Users accepted the idea of utilizing such a tool in their everyday lives.

During user-tests it was observed that some design and performance related issues had to be resolved in order to improve the system. Based on user feedback it can be derived that the following approaches will advance the system to better meet user needs.

An alternative approach to manipulating the tags is necessary. As this system used Flickr's relational method, in which various tags can be related in unorthodox/highly extended ways, the searches carried out with tags did not prove to be as efficient as the ones done with the semantic attributes. Hence it is necessary to tidy up the tags and produce new clusters and reconsider the relationship criteria among the tags.

The usability of the TagBall needs working on as well. To give the users a more flexible exploration experience, an easily accessible asset is necessary. A geometry that can be manipulated and elasticized will enable users to easily access the sought for tags.

The attributes need to be more comprehensive and enriched. In order to offer users an alternative access to finding a photograph, the alternative has to be rich in vocabulary just as the traditional way. The extraction of the underlying features can also be improved in order to enable experts to create more refined semantic attributes.

The system needs to be optimized for better performance and responsiveness. Reconsidering the technologies used within and more so the way how they were realized might be necessary to achieve a more robust system. A rapid system can be achieved by abandoning the current thin client and the application framework supporting it.

## 5.6 Summary

In this chapter, the user-centric tests carried out in order to evaluate the usability and functionality of $X^2$*Photo* were presented. The procedure followed during the test along with the survey conducted with a given questionnaire was described and the results were produced. When analyzed with the given feedback, the outcome of the tests showed interesting correlations between user habits and utilization of $X^2$*Photo* as well as a traditional tag-based image retrieval system used by Flickr. The strengths and weaknesses of the system were also observed and the steps that are required to be taken in order to improve it were outlined.

# Chapter 6

# Conclusion

## 6.1 Summary

This research investigated the possible benefits of augmenting conventional tag-based query techniques with subjective expert knowledge based on the manipulation of raw low-level data in exploring visual media. As related work suggested, in order to bridge the semantic gap the content of a photograph is not enough in inferring the semantics of a photograph. Hybrid image retrieval systems are considered to be encouraging developments in the field. Based on this notion we developed $X^2Photo$ which aimed to empower users in retrieving photographs from large collections using not only objective tags but also subjective expertise based on photographs' color space.

We collected 12000 photographs from Flickr along with their related tags. We considered the language of color theory to present a semantic space which could be utilized in order to search for photographs with a perceptive and aesthetic approach. The collected photographs were processed to extract their color spaces. We translated this data into high-level expressions that the end-user would use to pose verbal queries. A domain model was constructed and the semantic attributes were encoded into the system via SARA's authoring tool.

After experimenting with several prototypes we developed the final tool that presented a highly visual and attractive user interface which allowed end-users to interact with the system through certain access points. To make use of tags, the system offered the TagBall and the subjective expertise was presented through the AttBar. We introduced the concept of refining a search based on adjusting the attributes of a specific photograph.

The user test and the results of its accompanying survey showed how people originally communicate how a photograph is before what it portrays, but when utilizing a conventional tag-based system they tend to ignore the appreciative expressions as the system does not support such aesthetic relations. Hence this leads to limited searching via tagged simplifications of a photograph rather than the aesthetics of the photograph itself.

The types of photographs found by the users, using the more natural expressions offered by the system, indicate that this approach can be used to grant users the freedom they seek in relation to photograph searching. Injecting expert knowledge into a conventional system that only offers tag-based searching, allows users to freely express both the aesthetic of the photograph they want as well as the picture it conveys offering users an alternative pathway to access their large photograph collections.

## 6.2 Use Case

If we look at a leading image search engine, Google Images; we see that it has recently provided a few colors to be selected in order to have results with similar color spaces. Considering the innumerous amount of images they index and this new functionality they offer, it can be suggested that our idea can be integrated into their system when enhanced with different expert subscriptions. The new functionality wouldn't cause any overhead and users would still be able to pose verbal queries rather than choosing some basic colors.

## 6.3 Future Work

All the users agreed that the system was a powerful tool for exploring photographs and when asked if they could see a real-world application stemming from $X^2Photo$, all users concurred, as long as further improvements were made. Some engineering decisions will need to be reconsidered in order to offer a more robust system.

An alternative approach to manipulating the tags was found to be necessary. Rather than using Flickr's relational method, it is ideal to produce new clusters and reconsider the relationship criteria among the tags. We should also consider another geometry that can be manipulated and elasticized in order to enable users to easily access the sought for tags.

The attributes can also be enriched and the extraction of the underlying features can be improved with more sophisticated image analysis techniques in order to enable experts to create more refined semantic attributes. Furthermore as a specific expert vocabulary may not be suitable or correct for each individual, users should be able to subscribe to different experts to consider different perspectives. However within this approach ambiguities may arise due to the synonymous use of semantic attributes. To overcome this issue, an ontological structure of the words would be ideal and produce a finer hybrid system.

# Appendix A

# Evaluation Questionnaire

Do you consider yourself a photography enthusiast?

Do you have an account in an online photo sharing site such as Flickr, Picasa, etc.?

If so do you tag your photos and with what type of associations?

Do you have a personal blog in which you display photographs?

If so what kind of methods do you use to annotate them? (Tag them, use captures, titles, etc.)

When you need to find images, which image search engines or stock photography sites, or any other, do you use?

Please rate the following:

| | Very Good | Good | Poor | Very Poor |
|---|---|---|---|---|
| Discovery Space | | | | |
| TagBall | | | | |
| AttBar | | | | |
| Zoom | | | | |
| Extra Details | | | | |
| Focus | | | | |
| Favorites Area | | | | |
| Overall UI | | | | |

During exploration I found:

| | Very Useful | Useful | Not Useful | Completely Useless |
|---|---|---|---|---|
| Using Tags | | | | |
| Using Atts | | | | |
| Refining a Search based on focus image | | | | |
| Overall Search | | | | |

I found the system to be:

| | Strongly Agree | Agree | Disagree | Strongly Disagree |
|---|---|---|---|---|
| Attractive | | | | |
| Powerful | | | | |
| Empowering | | | | |
| Frustrating | | | | |
| Responsive | | | | |
| Slow | | | | |
| Extensive | | | | |
| Confusing | | | | |
| Straightforward | | | | |

Comments:

# Appendix B

# Domain Model


```
<?xml version="1.0" encoding="utf-8"?>
<DomainModel>
<Elements>
  <id>
    <location>
      <collection>rsp</collection>
      <hook>photo</hook>
      <node>id</node>
      <units></units>
      <returns>photo</returns>
    </location>
  </id>
  <title>
    <location>
      <collection>rsp</collection>
      <hook>photo</hook>
      <node>title</node>
      <units></units>
      <returns>photo</returns>
    </location>
  </title>
  <tag>
    <location>
      <collection>rsp</collection>
      <hook>photo</hook>
      <node>tag</node>
      <units></units>
      <returns>photo</returns>
    </location>
  </tag>
  <Hue>
    <location>
      <collection>rsp</collection>
      <hook>photo</hook>
      <node>Hue</node>
      <units></units>
      <returns>photo</returns>
    </location>
  </Hue>
  <Saturation>
    <location>
      <collection>rsp</collection>
```

```
      <hook>photo</hook>
      <node>Saturation</node>
      <units></units>
      <returns>photo</returns>
    </location>
  </Saturation>
  <Lightness>
    <location>
      <collection>rsp</collection>
      <hook>photo</hook>
      <node>Lightness</node>
      <units></units>
      <returns>photo</returns>
    </Lightness>
  </Lit>
  <Count>
    <location>
      <collection>rsp</collection>
      <hook>photo</hook>
      <node>Count</node>
      <units></units>
      <returns>photo</returns>
    </location>
  </Count>
  <exposureTime>
    <location>
      <collection>rsp</collection>
      <hook>photo</hook>
      <node>exposureTime</node>
      <units></units>
      <returns>photo</returns>
    </location>
  </exposureTime>
  </Elements>
  <SuperClasses>
    <SuperClass id="P" index="id">id</SuperClass>
  </SuperClasses>
</DomainModel>
```

Listing B.1: Simplified Domain Model

# Bibliography

[Addis et al., 2003] Addis, M., Boniface, M., Goodall, S., Grimwood, P., Kim, S., Lewis, P., Martinez, K., and Stevenson, A. (2003). Integrated image content and metadata search and retrieval across multiple databases. In *Image and Video Retrieval*, pages 91–100.

[Ames and Naaman, 2007] Ames, M. and Naaman, M. (2007). Why we tag: motivations for annotation in mobile and online media. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980, New York, NY, USA. ACM Press.

[AMFPHP, 2009] AMFPHP (2009). Flash remoting for php: A responsive client-server architecture. `http://www.amfphp.org/`.

[Bing, 2009] Bing (2009). Bing. `http://www.bing.com/`.

[Boutell and Luo, 2004] Boutell, M. and Luo, J. (2004). Photo classification by integrating image content and camera metadata. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, pages 901–904 Vol.4.

[Broder, 2002] Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, 36:3–10.

[Cai et al., 2004] Cai, D., He, X., Li, Z., Ma, W.-Y., and Wen, J.-R. (2004). Hierarchical clustering of www image search results using visual, textual and link information. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 952–959, New York, NY, USA. ACM.

[CameraFinder, 2009] CameraFinder (2009). Most popular brands. http://www.flickr.com/cameras/.

[Canon, 2009] Canon (2009). Picture styles. http://cpn.canon-europe.com/content/product/canon_software/colour_your_way.do.

[Champ, 2009] Champ, H. (2009). 3 billion. http://blog.flickr.net/en/2008/11/03/3-billion/.

[Cooliris, 2009] Cooliris (2009). Cooliris. http://www.cooliris.com/.

[Cui et al., 2008] Cui, J., Wen, F., and Tang, X. (2008). Real time google and live image search re-ranking. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, pages 729–732, New York, NY, USA. ACM.

[Datta et al., 2005] Datta, R., Li, J., and Wang, J. Z. (2005). Content-based image retrieval: approaches and trends of the new age. In *MIR '05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, pages 253–262, New York, NY, USA. ACM.

[Davis, 2000] Davis, S. (2000). *Color perception: Philosophical, Psychological, Artistic, and Computational Perspectives.* Oxford University Press.

[Eakins et al., 2004] Eakins, J. P., Briggs, P., and Burford, B. (2004). Image retrieval interfaces: A user perspective. In *Image and Video Retrieval*, pages 628–637.

[Enotic, 2009] Enotic (2009). Photology. http://www.getphotology.com/.

[Enser, 2000] Enser, P. (2000). Visual image retrieval: seeking the alliance of concept-based and content-based paradigms. *Journal of Information Science*, 26:199–210.

[Enser et al., 2006] Enser, P. G. B., Sandom, C. J., and Lewis, P. H. (2006). *Surveying the Reality of Semantic Image Retrieval.* Springer Berlin / Heidelberg.

[Exalead, 2009] Exalead (2009). Exalead. http://www.exalead.com/.

[eXistDB, 2009] eXistDB (2009). Open source native xml database. http://exist.sourceforge.net/.

[Fehrman and Fehrman, 2000] Fehrman, K. R. and Fehrman, C. (2000). *Color: The Secret Influence*. Prentice Hall.

[Flickner et al., 1995] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995). Query by image and video content: the qbic system. *Computer*, 28:23–32.

[Flickr, 2009] Flickr (2009). Flickr. http://www.flickr.com/.

[Frawley et al., 1992] Frawley, W. J., Shapiro, P. G., and Matheus, C. J. (1992). Knowledge discovery in databases - an overview. *Ai Magazine*, pages 57–70.

[Gage, 1999] Gage, J. (1999). *Color and Meaning: Art, Science, and Symbolism*. Berkeley; Los Angeles: University of California Press.

[Golder and Huberman, 2005] Golder, S. and Huberman, B. A. (2005). The structure of collaborative tagging systems.

[Gong, 1999] Gong, Y. (1999). Advancing content-based image retrieval by exploiting image color and region features. *Multimedia Syst.*, 7:449–457.

[Google, 2009] Google (2009). Google. http://www.google.com/.

[Gwap, 2009] Gwap (2009). Gwap. http://www.gwap.com/.

[Hampson, 2009] Hampson, C. (2009). Supporting personalised information exploration through subjective expert-created semantic attributes. In *3rd IEEE International Conference on Semantic Computing*.

[Hare et al., 2006] Hare, J. S., Lewis, P. H., Enser, P. G. B., and Sandom, C. J. (2006). Mind the gap: another look at the problem of the semantic gap in image retrieval. In Chang, E. Y., Hanjalic, A., and Sebe, N., editors, *Multimedia Content Analysis, Management, and Retrieval 2006*, volume 6073. SPIE.

[Hermitage, 2009] Hermitage (2009). Hermitage. http://www.hermitagemuseum.org/.

[Hoeber and Yang, 2006] Hoeber, O. and Yang, X. D. (2006). A comparative user study of web search interfaces: Hotmap, concept highlighter, and google. In *Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference on*, pages 866–874.

[Images, 2009] Images, G. (2009). Getty images. http://www.gettyimages.com/.

[imense, 2009] imense (2009). imense. http://www.imense.com/.

[iStockPhoto, 2009] iStockPhoto (2009). istockphoto. http://www.istockphoto.com/.

[Jackson, 2002] Jackson, R. S. (2002). *Wine tasting: a professional handbook*. Elsevier.

[Jau-Ling and Ling-Hwei, 2002] Jau-Ling, S. and Ling-Hwei, C. (2002). Color image retrieval based on primitives of color moments. *Recent Advances in Visual Information Systems*, pages 19–27.

[JEITA, 2002] JEITA (2002). Exif: Exchangeable image file format for digital still cameras: Exif version 2.2. Technical report, Japan Electronics and Information Technology Industries Association.

[Jing et al., 2005] Jing, F., Li, M., Zhang, H., and Zhang, B. (2005). A unified framework for image retrieval using keyword and visual features. *Image Processing, IEEE Transactions on*, pages 979–989.

[KDEG, 2009] KDEG (2009). Knowledge and data engineering group, trinity college, dublin. http://kdeg.cs.tcd.ie/.

[Kurgan and Musilek, 2006] Kurgan, L. A. and Musilek, P. E. T. R. (2006). A survey of knowledge discovery and data mining process models. *The Knowledge Engineering Review*, pages 1–24.

[Labeler, 2009] Labeler, G. I. (2009). Google image labeler. `http://images.google.com/imagelabeler/`.

[Lerman and Jones, 2006] Lerman, K. and Jones, L. (2006). Social browsing on flickr.

[Lew et al., 2006] Lew, M. S., Sebe, N., Djeraba, C., and Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.*, 2:1–19.

[Lewis et al., 2004] Lewis, P. H., Martinez, K., Abas, F. S., Fauzi, M. F. A., Chan, S. C. Y., Addis, M. J., Boniface, M. J., Grimwood, P., Stevenson, A., Lahanier, C., and Stevenson, J. (2004). An integrated content and metadata based retrieval system for art. *Image Processing, IEEE Transactions on*, 13:302–313.

[Marlow et al., 2006] Marlow, C., Naaman, M., Boyd, D., and Davis, M. (2006). Ht06, tagging paper, taxonomy, flickr, academic article, toread. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40, New York, NY, USA. ACM Press.

[Mehrabian and Valdez, 1994] Mehrabian, A. and Valdez, P. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, pages 394–409.

[Mislove et al., 2008] Mislove, A., Koppula, H. S., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. (2008). Growth of the flickr social network. In *WOSP '08: Proceedings of the first workshop on Online social networks*, pages 25–30, New York, NY, USA. ACM.

[Morton, 2004] Morton, J. (2004). *A Guide to Color Symbolism*. Color Voodoo.

[Nov et al., 2008] Nov, O., Naaman, M., and Ye, C. (2008). What drives content tagging: the case of photos on flickr. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1097–1100, New York, NY, USA. ACM.

[Parramon, 1989] Parramon, J. (1989). *Color Theory*. Watson-Guptill Publications, New York.

[PheXist, 2009] PheXist (2009). Php and perl classes for exist xml database. http://query-exist.sourceforge.net/.

[Picsearch, 2009] Picsearch (2009). Picsearch. http://www.picsearch.com/.

[Piximilar, 2009] Piximilar (2009). Piximilar. hhttp://ideeinc.com/products/piximilar/.

[Pixolu, 2009] Pixolu (2009). Pixolu. http://www.pixolu.de/.

[Pixsy, 2009] Pixsy (2009). Pixsy. http://www.pixsy.com/.

[Plataniotis and Venetsanopoulos, 2000] Plataniotis, K. N. and Venetsanopoulos, A. N. (2000). *Color Image Processing and Applications (Digital Signal Processing)*. Springer, 1 edition.

[PV3D, 2009] PV3D (2009). http://www.papervision3d.org/.

[Python, 2009] Python (2009). Python programming language – official website. http://www.python.org/.

[Rasiwasia et al., 2007] Rasiwasia, N., Moreno, P. J., and Vasconcelos, N. (2007). Bridging the gap: Query by semantic example. *Multimedia, IEEE Transactions on*, 9:923–938.

[Shen et al., 2000] Shen, H. T., Ooi, B. C., and Tan, K. L. (2000). Giving meanings to www images. In *MULTIMEDIA '00: Proceedings of the eighth ACM international conference on Multimedia*, pages 39–47, New York, NY, USA. ACM.

[Sinha and Jain, 2008a] Sinha, P. and Jain, R. (2008a). Classification and annotation of digital photos using optical context data. In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 309–318, New York, NY, USA. ACM.

[Sinha and Jain, 2008b] Sinha, P. and Jain, R. (2008b). Semantics in digital photos: A contenxtual analysis. *icsc*, 0:58–65.

[Smeulders et al., 2000] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22:1349–1380.

[Stüvel, 2009] Stüvel, S. A. (2009). Python flickr api interface. `http://www.stuvel.eu/projects/flickrapi`.

[Surowiecki, 2004] Surowiecki, J. (2004). *The wisdom of crowds : why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday.

[Tamura and Yokoya, 1984] Tamura, H. and Yokoya, N. (1984). Image database systems: A survey. *Pattern Recognition*, pages 29–43.

[Tuffield et al., 2006] Tuffield, M. M., Harris, S., Brewster, C., Gibbins, N., Ciravegna, F., Sleeman, D., Shadbolt, N. R., and Wilks, Y. (2006). Image annotation with photocopain. In *In Proceedings of Semantic Web Annotation of Multimedia (SWAMM-06) Workshop at the World Wide Web Conference 06. WWW*, pages 22–26.

[Uchihashi and Kanade, 2005] Uchihashi, S. and Kanade, T. (2005). Content-free image retrieval by combinations of keywords and user feedbacks. *Image and Video Retrieval*, pages 650–659.

[Villa et al., 2008] Villa, R., Gildea, N., and Jose, J. M. (2008). A faceted interface for multimedia search. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 775–776, New York, NY, USA. ACM.

[von Ahn and Dabbish, 2004] von Ahn, L. and Dabbish, L. (2004). Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, New York, NY, USA. ACM.

[W3C, 2009] W3C (2009). W3c. http://www.w3.org/.

[WP-Cumulus, 2009] WP-Cumulus (2009). Wp-cumulus. http://wordpress.org/extend/plugins/wp-cumulus/.

[Xu et al., 2009] Xu, S., Jin, T., and Lau, F. C. M. (2009). A new visual search interface for web browsing. In *WSDM '09: Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 152–161, New York, NY, USA. ACM.

[Yahoo!, 2009] Yahoo! (2009). Yahoo! http://www.yahoo.com/.

[Yu et al., 2002] Yu, H., Li, M., Zhang, H.-J., and Feng, J. (2002). Color texture moments for content-based image retrieval. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 3, pages 929–932 vol.3.

Sir Galahad: Is there someone else up there we can talk to?
French Soldier: No, now go away or I shall taunt you a second time.
*Monty Python and the Holy Grail 1975*