

Personal Photo Annotation

Nataliya Alexander

September 9, 2005

ABSTRACT

Personal digital photographic collections grow quickly in size. Managing and annotating these collections becomes difficult and requires much time and effort. This research investigates into the ways people annotate their personal photographic collections to gain a deeper understanding of photo annotations in order to help people manage their digital personal photo collections. The research also highlights the differences that exist in annotations by males and females and the attributes that describe these differences. Various attributes of annotations are studied. They consist of length, structure, word classes of the English language, whether an annotation is artistic and entropy as a measure of information of annotation. The findings include the inverse relation between the percentage of proper nouns and verbs in annotations, the increase in emotions in descriptions of animate and inanimate objects, the decrease in emotions in descriptions of people, the presence of more emotions in annotations by females and the discovery of four groups of annotations each with different combination of attribute values. The findings highlight the value of related work in automatic annotations because 67% of supplied annotations are merely descriptions. However, the remaining 33% contain emotions expressed by the author and these emotions cannot be captured through automatic annotation.

CONTENTS

1. <i>Introduction</i>	3
2. <i>Technical Background</i>	5
2.1 EXIF (Exchangeable Image File Format) Data	5
2.2 MPEG-7 (Moving Picture Experts Group) standard	6
3. <i>Related Work</i>	7
3.1 Text annotation	7
3.2 Geo-referencing and naming	8
3.3 Unsupervised learning models for automatic annotation	9
3.4 Managing and annotating digital photo collections on handheld devices	11
3.5 Photographic interpretation with relation to tourism and visual anthropology	12
4. <i>Tools</i>	15
4.1 EXIF Class Library	15
4.2 Web application	22
5. <i>Experiment</i>	31
6. <i>Analysis</i>	33
6.1 Data familiarisation, attribute enrichment and cleansing	33
6.2 Data Visualisation	35
6.3 Clustering	44
6.4 Analysis of Clusters	47
7. <i>Principle Component Analysis (PCA)</i>	52
8. <i>Conclusion</i>	54

<i>Appendix</i>

62

1. INTRODUCTION

Photographs enhance our memories. A photograph can mean different things to different people. This can be due to the surroundings or immediate environment of the viewer or to their cultural background. The photographic annotation helps to interpret the meaning of a photograph, adds more detail to the image and helps to remind the owner of events, locations and people.

In times past there was a distinct connection between one's means and one's records as only the very wealthy could afford to have portraits and paintings. From the inception of the photographic image to current digital photography, the ability to keep annotated visual records has spread across the social spectrum.

The falling cost of digital camera technology makes it more affordable to the wider population. In 2003 digital camera sales overtook film-based camera (Gussow 2004). Digital photography changes the way people use personal images, as the cost of each image is effectively zero.

The task of organising and managing a photographic collection becomes more difficult as the collection increases in size. Locating a photograph in a collection constitutes much effort and is a time consuming activity, particularly for collections containing thousands of photographs. Annotation of a large collection of digital photographs can be boring and tedious. At the time of annotation it is difficult to see the future benefits.

Personal Photo Annotation research aims to obtain an understanding of how people annotate their personal photographs and the types of personal photographs people take. The research also attempts to establish how people of different ages, social and ethnic backgrounds annotate their personal photographs. A deeper insight into personal photo annotations will help us to gain better understanding of photo annotations and bring us closer to being able to help people manage their digital personal photo collections.

The remainder of this thesis is structured as follows:

There are 7 sections. Section 2 contains technical background information. The

related work is described in Section 3. Section 4 includes the description of tools. Section 5 describes the experiment and the quality and quantity of data obtained. Section 6 reports on methods used to analyse annotation data and documents the analysis. Section 7 contains the results of Principle Component Analysis (PCA) to verify some findings in the analysis in Section 6. Lastly, Section 8 summarises observations, relates the findings to the research in the related work and delivers conclusions.

2. TECHNICAL BACKGROUND

2.1 *EXIF (Exchangeable Image File Format) Data*

Image files recorded using digital cameras contain EXIF header embedded in the image file. EXIF is an exchangeable image file format for digital still cameras. It is developed by JEITA (Japan Electronics and Information Technology Industries Association) and specifies formats for images, sound and tags for digital still cameras (Exif 2005). EXIF information includes various camera settings and attributes that describe primary and thumbnail images.

Some cameras have an interface to connect to a GPS (Global Positioning System) unit (Kodak 2004; Nikon 2004) or utilize a CompactFlash WAAS GPS card slot (GeoSpatial Experts 2004). GPS allows the recording of the exact latitude and longitude position the photograph was taken.

There are two parts of information in the EXIF header that deserve special attention. These are the timestamp and the location information.

The timestamp details can be used to extract metadata such as season (assuming that the hemisphere the photo was taken in is known), part of day: morning, afternoon, evening (assuming that the time zone the photo was taken in is known), month, century, and year. The timestamp information can be linked to calendar events such as holidays. Although, accurate calendar information is required as different cultures and religions mark events that are specific to them.

Location details expand the timestamp metadata collection by introducing new items such as country, continent, city and town. Many more items can be discovered (Naaman et al. 2004d) using timestamp and location information and include light status, time zone, temperature, and weather status.

2.2 MPEG-7 (*Moving Picture Experts Group*) standard

MPEG-7 standard is formally known as Multimedia Content Description Interface. The standard presents an interoperable solution that allows indexing, searching and retrieval of audio-visual resources. The main objective of the standard is to provide a uniform way of describing information about the content of an audio-visual resource.

The standard implements four elements used to describe multimedia content: descriptors (D), description schemes (DS), description definition language (DDL) and coding schemes.

A Descriptor is a structure written in XML. The Descriptor specifies various features of the audio-visual content such as time, location, colour, and texture. It can either embody another descriptor or values. Description Schemes use XML Schema Language to outline relationships between descriptors. Description Definition Language specifies the syntax for the Descriptors and Description Schemes.

MPEG-7 is a complex standard. However, it provides a very flexible and exhaustive way to describe multimedia content.

MPEG-7 standard serves as a base for the Caliph and Emir prototypes (Lux et al. 2004). The prototypes are created in an attempt to develop suitable tools that express additional semantics of a photograph and provide a mechanism for semantic annotation and retrieval of the photographs.

3. RELATED WORK

3.1 Text annotation

Textual photo annotation goes back over one hundred years ago to the times when the first photographs were taken. Annotation helps to add extra details to the images and revive memories especially when sharing the photographs with others.

Annotation can also be compared with a very short story. Often even a short sentence is sufficient to satisfy viewer's curiosity and to understand what happened. It is due to the fact that terms used in the annotation text are closely relevant to the image content. This property has also been observed in a collection of news photographs (Edwards et al. 2003).

Usually people organise their photographs in chronological order. An annotated collection of photographs produces a story line about someone's activities for a certain period of time. The storytelling with digital photographs is investigated further in the study where a portable device is used in place of a traditional photo album (Balabanovic et al. 2000). The observations report spontaneous interpersonal interaction. The authors believe that viewing, creating and telling stories activities are interrelated. Therefore, the modeless interfaces provide the best mechanism for such interaction.

Digital photo collections grow quickly in size. It takes time and effort to annotate each photo. On many occasions a batch of photographs depicts the same thing, for instance, a dog. Captions often will only be added to the best photographs.

There are a number of studies that research the problem of digital photo annotation and propose various solutions. These studies can be summarised under the following two categories: automatic annotation using timestamp and location information, automatic annotation based on unsupervised learning.

3.2 Geo-referencing and naming

Geographical coordinates and timestamp are a valuable source of information because they encompass data that can be interpreted into memories of events and visited locations. Often events and locations are interlinked, for instance, a honeymoon.

The dependencies found between locations and events have been exploited further (Naaman et al. 2004c) to produce a PhotoCompas system. PhotoCompas presents a solution that automatically organises a collection of digital photographs and annotates them with geographical names. The system design is based on an algorithm that detects events and accounts for changes in locations.

The authors observed that people take photographs in bursts. Different events and locations are discovered through comparison of the time gap and the geographical distance between the images. The output of the algorithm can be described as a three-step process: discover and define *event clusters*; then discover and define *location clusters*; and refine *event clusters*.

The photographs are grouped into meaningful clusters and textual geographical names are assigned to the clusters and the photographs.

The beta testing of the PhotoCompas system confirms that users find the organization of the photographs very useful. Also, the textual descriptions given to the photographs and the clusters are very similar to those the users of the system would have given themselves.

A related study (Naaman et al. 2004a) compares PhotoCompas interface and the World Wide Media Exchange (WWMX) application designed to manage digital photographs.

Both systems offer two-dimensional interaction - in location and in time - with digital photo collections. The WWMX utilizes a powerful map-based and timeline interface. A user can search for photographs navigating through a timeline and to the location on a map.

The PhotoCompas system has a higher score in perceived helpfulness for searching and browsing of photographs in time. However, WWMX interface is more efficient for searching and browsing of photographs in locations. The results of subjective user evaluation also show that WWMX offers more entertaining and satisfying experience.

The location information is also useful in collaborative photo annotation. It is illustrated in the LOCALE system (Naaman et al. 2004b). The system enables a term search over unlabelled photo collections and proposes automatically generated labels for unlabelled photographs. The system utilizes a central server with a database. Users contribute to the database by submitting photographs that are labelled and location stamped. These labels are used to search unlabelled photo collections and annotate unlabelled photographs. This saves time and effort taken to annotate photographs and creates a process of collaboration.

3.3 *Unsupervised learning models for automatic annotation*

Significant research has been devoted to automatic image annotation based on unsupervised learning (Barnard et al. 2003a; Duygulu et al. 2002; Edwards et al. 2003; Barnard et al. 2003b; Forsyth 2001; Barnard et al. 2003c, 2001). This research advances on two observations. The first observation is that separately text and images can be very ambiguous in meaning but when used in combination resolve ambiguity in most cases. The second observation is that the parts that are visually explicit are omitted from captions and, instead, captions normally contain the parts that are hard to deduce visually, just by looking at an image. This property is exploited further to create models that are suitable for browsing of photo collections, effective search, automatic annotation, and object recognition.

Furthermore, a task of predicting text for images is categorised into annotation and correspondence (Barnard et al. 2003a). For the task of annotation a whole image is used to predict text. For the task of correspondence an image is segmented into regions using a set of image features and computer vision techniques. Each region represents a tangible object such as sun, clouds and face. Then, unsupervised learning is applied as a hierarchical combination of symmetric and asymmetric clustering (Barnard et al. 2001) in order to establish association of a particular object to a specific word. Asymmetric clustering links images to clusters. Symmetric clustering links images and their features. The result is a binary tree. The path from the root of the tree to the leaf node represents a cluster. Each node on the path contains a probability of an image region and a word in that node. The probability that an image belongs to a cluster is calculated as a sum over

probabilities of image regions and words present in the nodes of a cluster weighted by the probability that an image belongs to the cluster. Note that the word and the region do not have to correspond to the same node but are only required to come from the same cluster.

Clusters may share nodes because the same image region or a word can be used to describe many images. Nodes at the top of the hierarchy contain general terms and image regions. Nodes located closer to the leaf node contain image regions and words that are more specific to a particular image and occur only a few times in a collection. This organisational structure of nodes illuminates the relation of nodes and topics because images that share nodes are likely to belong to the same topic. Furthermore, a node represents a general topic if positioned at the top of the tree, and a more specific topic if located closer to the leaf node. This structure can be used for constructing a useful browsing mechanism for a collection of digital images.

The clustering model is used further in this research to create an alternative model (Barnard et al. 2001). The alternative model differs from the clustering model because it introduces a constraint for words and regions. The constraint requires a word and a region to come from the same node in a cluster. This strengthens a link between words and regions. The alternative model offers higher precision for a search query.

The study extends further to derive three models for automatic image annotation: multi-modal hierarchical aspect model, mixture of multi-modal Latent Dirichlet allocation, and simple correspondence model (Barnard et al. 2003a). Additionally, several variations of each model are implemented and used in experiments.

The authors point out that it is significantly harder to evaluate the performance of the models for the correspondence task in comparison with the annotation task. This is due to the fact that large datasets of labelled image regions simply do not exist. The only way to measure performance for the correspondence task is to view image on an individual basis. This task has been accomplished by selecting and verifying correspondence in 100 images (Duygulu et al. 2002).

The results of the experiments with three models suggest that correspondence task is useful for annotation. The model that proved to be the most fruitful is an integrated model of hierarchical clustering and simple correspondence (Barnard et al. 2003a).

The problem of automatic annotation is also regarded as a machine translation prob-

lem (Barnard et al. 2003b) and analogous to learning a lexicon (Duygulu et al. 2002). The difference is that instead of translating text from one language to another, translation occurs between image regions and words of annotation text. The lexicon consists of a vocabulary of words present in annotation texts of images. Image is segmented into regions with region features vector quantised. Each region is also referred to as a blob. Initially, the probability of a word given a blob is estimated and recorded in a table. Then, the co-occurrence of a word and a blob is used to refine the probability table. During testing stage, the words with the maximum probability given a blob are used to construct annotation text for an image. Changing the threshold value of the maximum probability produces better results because the probability of rare words is shifted toward boosting the appearance of more commons words. A usefulness of digital image management systems has also been addressed (Forsyth 2001). The author's opinion is that there is a big gap between user needs and what image retrieval systems can currently offer. The author claims that too many emphases are placed on search. More work is required to understand user needs such as how a user manages image collections, what makes an image collection management system useful to a user and how an image collection should be structured in order to make it useful and meaningful for browsing. Another observation of paramount importance is that annotated images prove to be valuable in practice because the combination of image and annotation resolves ambiguity and provides a background for establishing a topic. It has also been mentioned that new users prefer to browse a collection while users that are familiar with the collection tend to use its search functionality (Forsyth 2001).

3.4 Managing and annotating digital photo collections on handheld devices

Personal digital photo collections consisting of hundreds or even thousand of photographs can be stored on a single PDA device. There are times when a user would like to share photographs with others. These photographs may be taken several months ago and are mixed with more recent ones. Unless there is a defined structure of folders to organize large personal photo collections, the task of finding the photographs can be difficult.

In addition to digital cameras, photographs can also be taken using camera phones

and PDAs. With mobile phone technology location information is always available to a user. A research (Wilhelm et al. 2004) into photo annotation on a camera phone suggests that the immediate availability of time and location information allows users to annotate photographs at the time of capture. Furthermore, the same research proposes a collaboration system that allows users to reuse descriptions of others stored in a central location in order to save time and effort.

The study also highlights *The Power of Now* as another aspect of camera phone technology. It is reported that users take more interesting and unique pictures using camera phones.

One study proposes a photo browser to support large personal digital photo collections on PDAs (Harada et al. 2004). The study also compares automatic and manual organization of photos.

Two different browser interfaces are suggested for automatic organization of photos: Baseline and Timeline browser. Both browsers implement time clustering algorithm on the assumption that a user takes photographs in bursts. The clustering algorithm partitions the photo collection into meaningful clusters presented to the user in a form of albums.

The Baseline browser implements a folder-based interface. Timeline browser interface is split into three columns. Two columns of pictures on the right and on the left hand side represent albums of photos grouped into major clusters. The column in the middle is a timeline scrollbar split into month sections. The Timeline scrollbar can be used to quickly navigate to clusters of photos taken within a certain period of time.

The results of the experiments report that the automatic organization of photograph performs almost as well as manual organization for search and browsing tasks. In addition, searching and browsing tasks take less time to complete in the Timeline browser than in the Baseline browser after a user becomes more familiar with both interfaces.

3.5 *Photographic interpretation with relation to tourism and visual anthropology*

Different people from all countries will either take pictures of the landmarks for their own sake and intrinsic value or mementos with family member with scenic backgrounds.

The presence of the traveller in the photograph shows a clear association between the location and the person. This association is very important because it holds a connection between person and location in traveller's memory. The connection subconsciously revives the emotions that a traveller experienced during the time spent in that location.

A collection of essays (Crouch and Lubgren 2003) dedicated to the visual culture and tourism, provides a deeper view into connections and interaction between photography and tourism.

In the essay about the holidaymakers who visited the Isle of Man in 1950s (Sandle 2003) a photograph is represented as a testimony and reminiscence. The testimony of a successful holiday is expressed through inclusion of the holidaymaker himself in the photograph, often with a group of friends and family and even a landlady. There is an additional meaning in the group photography, a social significance of the holidaymaker, evidence of successful interaction with other people known to the holidaymaker or complete strangers.

There are other roles of photography in tourism (Deriu 2003): relationships between what was seen by the traveller are preserved in the photograph; photographs restore forgotten memories of visited places and people met during the journey; travelling is a way to increase photographic collection; and, the photograph is evidence of the reality captured by a camera mechanism, a reality that is never questioned as being false.

One study suggests (Liechti and Ichikawa 1999) that digital photographs play a significant role in maintaining social awareness and interaction. The simple act of receiving and sending messages and photographs creates a connection between the sender and the recipient. The photographs emphasize this connection and perhaps make it more tangible. The proposed framework to capture, annotate and distribute photographs is based around intelligent devices (Liechti and Ichikawa 2000) such as fridge panel that has a built in display device to view photographs.

In Visual Anthropology (Zeitlyn 2003) photographs can be used in two distinct ways as well as their combination: a photograph as a subject of the anthropological study; and, a photograph as a tool for gathering visual material to assist an anthropologist's research.

A study (Scherer 1992) into photographic documents used in an anthropological enquiry suggests that the meaning of the photograph is obtained by combining the viewer's

interpretation, understanding the photographer's intention and the photograph itself as an artefact. The author of the study regards the social interaction between the photographer, the viewer and the subject of utmost importance that determines the sociocultural meaning of the photograph.

Contents of a photo annotation depend on how a viewer interprets the photograph. Thus, it is important to establish what affects viewer's interpretation. The suggestion is that an interpretation of the photograph depends on what we see on the photograph. In one study (Berger 1972) the author claims that the viewer's beliefs and knowledge form the way of seeing. What is seen in the photograph may also be affected by one's ability to imagine or fantasize. Due to this distinction, two types of photo annotations can be observed: creative captions and simple descriptions.

Interpretation of a photograph can also be influenced by whether a viewer attempts to understand photographer's intention. This adds an additional meaning to the photograph, perhaps something a viewer would not see otherwise. One suggestion is that there are two factors that govern this: whether the photograph is personal; and whether the person interpreting the photograph was somehow involved in the process of taking the photograph.

Personal photographs are more interesting because there is proximity between the viewer and the photograph. The proximity is often expressed through personal contact or awareness between the viewer and the contents of the photograph.

The interpretation of a photograph is also influenced by time. People perceive past events differently as time goes by. Moments that were previously considered unimportant may now be viewed as a turning point in one's life.

4. TOOLS

Two tools have been designed for the purpose of the experiment. These tools include EXIF class library and the Collector web application. Detailed design and functionality of the tools is outlined below.

4.1 *EXIF Class Library*

EXIF information is embedded in a JPEG file and includes descriptions of image and digital camera information and a thumbnail image. It is recorded in compliance with the JPEG DCT format (JEITA 2002). EXIF information is useful for comparing images in order to establish what settings should be used under certain conditions to produce the best results. It also enables recording of user comments and image description to a JPEG file. Embedding information into a JPEG file means that it can be distributed to other users with no extra requirements to store it in a separate file. Users can view EXIF information using any software that supports the reading of EXIF data. EXIF information is also embedded in the original TIFF files.

EXIF library is a java class library designed to read and write EXIF information to a JPEG file. It complies with JEITA CP-3451(JEITA 2002) standard. Firstly, we discuss the structure of a JPEG file and then describe functionalities of the class library. We adopt hexadecimal representation for values used in the description of EXIF class library.

JPEG File Interchange Format is compression standard for images. JPEG or JPG files are compressed image files created according to JPEG File Interchange Format standard (Hamilton 1992). The structure of a compressed JPEG image file is shown in Figure 4.1.

Every JPEG file must start with '0xFFD8' and end with '0xFFD9'. These values are referred to as SOI (start of image) and EOI (end of image) accordingly. There can also be several **markers** embedded in a JPEG file. Each marker holds a chunk of information and starts with '0xFFXX' value, where XX is its number. SOI and EOI are special types

SOI	Start of Image
APP1	Application Marker Segment 1 (Exif Attribute Information)
(APP2)	Application Marker Segment 2 (FlashPix Extension data)
DQT	Quantization Table
DHT	Huffman Table
(DRI)	(Restart Interval)
SOF	Frame Header
SOS	Scan Header
	Compressed Data
EOI	End of Image

Fig. 4.1: Structure of compressed file (JEITA 2002).

APP 1 Marker
APP1 Length
Exif Identifier Code
TIFF Header
0 th IFD
0 th IFD Value
1 st IFD
1 st IFD Value
1 st IFD Image Data

Fig. 4.2: Structure of APP1 Marker (JEITA 2002).

of marker because they do not carry any data (Tachibanaya 2001).

Marker '0xFFE1' is an Application Marker 1 (APP1 Marker). Figure 4.1 shows presence of APP1 Marker in a compressed image file. APP1 Marker contains EXIF attribute information and is used to store EXIF tags. Its structure is presented in Figure 4.2. APP1 Marker and its tags are vital in the design of EXIF class library. Detailed description of APP1 Marker is presented below.

APP1 Marker starts with '0xFFE1', followed by 2 bytes that hold APP1 Marker data size. Total data size of APP1 marker must not exceed 64Kbytes or '0xFFFF'. As per Figure 4.2, after APP1 Marker data size bytes EXIF Identifier Code bytes follow. EXIF Identifier Code bytes are also referred to as EXIF header.

The values of the EXIF header bytes must be '0x457869660000', where '0x45786966' bytes are used to represent ASCII character string "Exif" and '0x0000' are 2 bytes used as null termination characters. The presence of the EXIF header in application marker means that this marker is an EXIF marker. EXIF header is followed by TIFF Header.

TIFF Header contains information about byte order used to encode tags. Next 4 bytes recorded after TIFF header contain offset value to the 0th Image File Directory (IFD). Note that all offset values used in APP1 Marker are recorded in relation to the first byte of TIFF Header. This means that all offset values recorded in IFDs and tags are calculated from the first byte of TIFF header.

The value of byte order bytes is '0x4D4D' when "Big Endian" byte order is used and represents ASCII string "MM". "MM" stands for Motorola. The value of byte order bytes is '0x4949' when "Little Endian" byte order is used and represents ASCII string "II". "II" stands for Intel.

The APP1 Marker also consists of chunks of information known as Image File Directories (IFDs). Figure 4.2 shows two main IFDs: 0th IFD and 1st IFD. The contents of every IFD comprises of tags. IFD tag entries are followed by 4 bytes that contain either offset value to the next IFD or are filled with null values '0x00'. An offset value is used to calculate the address of the next IFD as the address value of the first TIFF header plus offset value.

There are also several other IFDs. These IFDs are Exif SubIFD, GPS IFD and Interoperability IFD. But next IFD offset values to their addresses are recorded in tags that are specially allocated to. 0th IFD contains the *Exif IFD Pointer* tag used to hold offset value Exif Sub IFD. 0th IFD may also contain the *GPS IFD Pointer* tag used to hold offset value to GPS IFD. The presence of the *GPS IFD Pointer* in 0th IFD depends on whether any GPS information has been recorded. Exif SubIFD may contain the *Interoperability IFD Pointer* tag used to hold offset value to Interoperability IFD.

In total, there can be the maximum of five IFDs in APP1 Marker. For 0th IFD next IFD offset bytes indicate offset to the next main IFD, the 1st IFD. For the 1st IFD, next IFD offset bytes are filled with null values because this is the last IFD in APP1 Marker. For Exif SubIFD, Interoperability IFD and GPS IFD next IFD offset bytes are filled with nulls. Figure 4.3 shows a structure of a typical APP1 Marker. Figure 4.4 shows a structure of a typical APP1 Marker with GPS information.

IFD information is recorded in the form of tags. 0th IFD describes primary image and contains tags such as orientation, colour space and resolution unit. Exif SubIFD contains digital camera information such as flash, ISO speed ratings and lens focal length.

APP1 Marker	
APP1 Data Size	
Exif Header	
TIFF Header	
0th IFD	
	Number of tag entries
	0 th IFD tags - primary image data tags - Exif SubIFD Pointer tags
	1 st IFD Offset
0th IFD Data for tags with values longer than 4 bytes	
Exif Sub IFD	
	Number of tag entries
	Exif SubIFD tags - digital camera data tags - Interoperability IFD Pointer tag
	Offset to next IFD (set to 0x00000000 in Hex)
Exif SubIFD Data for tags with values longer than 4 bytes	
Interoperability IFD	
	Number of tag entries
	Interoperability IFD tags
	Offset to next IFD (set to 0x00000000 in Hex)
1st IFD	
	Number of tag entries
	1 st IFD tags - thumbnail image data tags
	Offset to next IFD (set to 0x00000000 in Hex)
1st IFD Data for tags with values longer than 4 bytes	
Thumbnail Image Data	

Fig. 4.3: A structure of a typical APP1 Marker.

APP1 Marker	
APP1 Data Size	
Exif Header	
TIFF Header	
0th IFD	
	Number of tag entries
	0 th IFD tags - primary image data tags - Exif SubIFD Pointer tag - GPS IFD Pointer tag
	1 st IFD Offset
0th IFD Data for tags with values longer than 4 bytes	
Exif Sub IFD	
	Number of tag entries
	Exif SubIFD tags - digital camera data tags - Interoperability IFD Pointer tag
	Offset to next IFD (set to 0x00000000 in Hex)
Exif SubIFD Data for tags with values longer than 4 bytes	
Interoperability IFD	
	Number of tag entries
	Interoperability IFD tags
	Offset to next IFD (set to 0x00000000 in Hex)
GPS IFD	
	Number of tag entries
	GPS IFD tags
	Offset to next IFD (set to 0x00000000 in Hex)
1st IFD	
	Number of tag entries
	1 st IFD tags - thumbnail image data tags
	Offset to next IFD (set to 0x00000000 in Hex)
1st IFD Data for tags with values longer than 4 bytes	
Thumbnail Image Data	

Fig. 4.4: A structure of a typical APP1 Marker with GPS information.

Tag Code	Data Format	Number of Components	Value or Offset Bytes
2 bytes	2 bytes	4 bytes	4 bytes
Example of <i>Orientation</i> tag:			
0x0112	0x0003	0x00000001	0x00000001

Fig. 4.5: A template of a tag structure. Includes example of *Orientation* tag.

Data Format Value		Name	Meaning	Bytes Per Component
decimal	hex			
1	0x0001	Unsigned Byte	8 - bit unsigned integer	1
2	0x0002	ASCII String	8 - bit byte containing 7 - bit ASCII code	1
3	0x0003	Unsigned Short	16 - bit unsigned integer	2
4	0x0004	Unsigned Integer	32 - bit unsigned integer	4
5	0x0005	Unsigned Rational	The first 32 - bit as unsigned integer, represent numerator. The remaining 32 - bit as unsigned integer, represent denominator.	8
7	0x0007	Undefined	8 - bit byte that can take any value depending on its implementation.	1
9	0x0009	Signed Integer	32 - bit signed integer	4
10	0x000A	Signed Rational	The first 32 - bit as unsigned integer, represent numerator. The remaining 32 - bit as unsigned integer, represent denominator.	8

Fig. 4.6: Details and description of tag data formats.

Interoperability IFD contains only two pieces of information, interoperability index and interoperability version. GPS IFD contains GPS information such as latitude, longitude, GPS time (atomic clock) and altitude. Lastly, 1st IFD contains information about the thumbnail image.

There are five levels of tag support: mandatory, optional, recommended, not recorded and included in JPEG marker and so not recorded. A complete list of tags for each IFD and their support level can be found in JEITA standard (JEITA 2002). Usually, 0th IFD is recorded immediately after TIFF header. In such case its offset value is '0x00000008' and equals to 8 in decimal.

As mentioned earlier, each IFD consists of tags. A tag is a structure that holds a piece of information that describes a single attribute. This also makes a tag the smallest piece of meaningful information found in APP1 Marker. Tag size must be exactly 12 bytes. A template of a tag structure is presented in Figure 4.5. There are 8 tag data formats. Figure 4.6 contains details for each data format. According to JEITA CP-3451 standard some tags must have a fixed component count while others can be of any length. Tags that belong to ASCII string or Undefined data formats usually can have any number of

Character Code	Character Code Bytes (in hex)	References
ASCII	0x41, 0x53, 0x43, 0x49, 0x49, 0x00, 0x00, 0x00	ITU-T T.50 IA5
JIS	0x4A, 0x49, 0x53, 0x00, 0x00, 0x00, 0x00, 0x00	JIS X208-1990
Unicode	0x55, 0x4E, 0x49, 0x43, 0x4F, 0x44, 0x45, 0x00	Unicode Standard
Undefined	0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00	Undefined

Fig. 4.7: Descriptions and details of character codes used in *User Comment* tag (JEITA 2002).

components.

The last 4 bytes in a tag structure are used to store either the actual value of a tag or the offset bytes to the value if total data length of the value exceeds 4 bytes. For instance, if a tag *Model* stored in 1st IFD has a value “Powershot A300” than its total data length equals to 14 bytes (14 ASCII characters as 14 components times ASCII string bytes per component 4.6) and the value bytes contain the offset bytes. The actual value is stored at the address calculated as the address of the first byte of TIFF header plus the value of the offset bytes. When designing an EXIF writer, it is important to make sure that if any tag value has been modified then the offset bytes values for all IFDs and tags are updated, and updated values are calculated in relation to TIFF Header. Another important point is that ASCII string and Undefined data format bytes are always stored using “Big Endian” byte order and, thus, are not affected by the byte order specified in TIFF Header.

All tags share common properties such as tag structure, data format, support level and location in IFD but some have a very specific implementation. A example of such implementation is *User Comment* tag. *User Comment* is recorded using Undefined data format. This means that it can store any type of data and the ability to read and write data correctly is part of its specification.

For *User Comment* tag the first 8 bytes are used to identify specify a character code. Figure 4.7 contains descriptions of permitted character codes and corresponding bytes. Actual user comment bytes must follow after 8 bytes of character code. *User Comment* tag supports unicode encoded character string. This means that *User Comment* tag can be used to store comments in languages such as Chinese, Arabic and Japanese that require 16 bits per character.

JEITA standard also defines a special tag *MakerNote*. This tag belongs to an Undefined data format and contains other tags inside it. Tags that reside inside *MakerNote* tag are specific to a particular camera manufacturer. *MakerNote* allows different camera manufacturers record additional information not specified by standard tags. EXIF class

library can identify, read and write *MakerNote* tag but cannot decode it.

Common and specific tag properties are implemented in EXIF class library through inheritance. **Tag** class defines common tag properties and behaviour such as tag code, tag description, **toString** method and **getMeaning** method. Every tag described in JEITA specification has been implemented as a separate class that inherits from **Tag** class. In total, there are 121 tag classes.

There is also a package **util**. It contains the following classes: 8 classes that implement tag data formats; **ByteConverter** (Thang 2005) class that uses data formats to convert bytes to values and values to bytes; and **TagType** class which is a base class for tag data format classes.

EXIF class library also contains **IFD** class. **IFD** class implements properties and behaviour necessary to read and write IFDs to APP1 Marker. A **Tag** object can be added to or removed from IFD using methods **addTag** and **removeTag** accordingly. If the tag's value has been modified then it is possible that the size of the tag data has changed. The consequences of this action include: updating value bytes of the tag to either the actual value or the value of offset bytes; updating number of components value of the tag; and re-calculating offset byte values for all tags and IFDs. This means that not only the updated tag is affected but also all tags in all IFDs, the size of the IFD where the tag resides and the size of the APP1 Marker are also affected. **IFD** class holds all tags that exist in that IFD in a **Vector** object. When update happens only **setValue** of **Tag** is updated. Offset bytes of each tag, IFD directory size where that tag resides, offset for all IFDs and APP1 Marker size are all updated during the stage of compiling a byte stream for APP1 Marker.

Total APP1 Marker size should not exceed 64 K bytes. EXIF class library does not implement validation mechanism that informs a user if modification to a particular tag breaches this constraint. However, the total size of the APP1 Marker is calculated before it is written to a file and provides a suitable place where this validation mechanism can be implemented.

Other important classes include **TagAnalyser** and **ImageAnalyser**. Class **TagAnalyser** is a utility class. It is used to instantiate a correct object of one of 121 tag classes for a particular tag given its tag code. Class **ImageAnalyser** is responsible for reading and writing tags, creating appropriate IFD and **Tag** objects, and calling appropriate methods

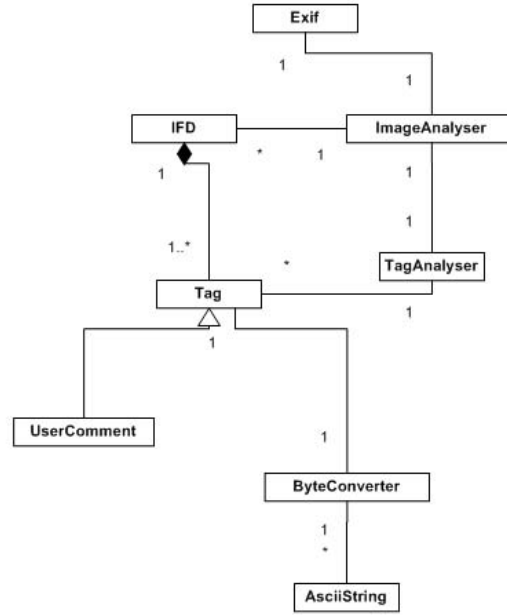


Fig. 4.8: EXIF Library UML diagram.

to update offset values. This class also contains method `getAllTags` that returns a `Vector` object populated with `Tag` objects extracted from APP1 Marker of a JPEG file. The EXIF class library can be extended to output EXIF information to any type of file such as XML or CSV.

The EXIF class library also provides class `Exif` as a point of entry to `ImageAnalyser` functionality. `Exif` class enables quick access to methods for reading and writing user comments. For the purpose of this project, it is only required to write and modify `UserComment` tag. Hence, for writing tags only `UserComment` tag is implemented.

The EXIF class library provides sufficient functionality for the purpose of this project and can be extended further to accommodate updating tags other than `UserComment`.

The EXIF class library UML diagram is shown in Figure 4.8. The diagram does not accommodate all 121 tag classes but, instead, shows only `UserComment` class as an example. The diagram also does not accommodate all 8 data format classes but, instead, shows only `AsciiString` class as an example.

4.2 Web application

The Collector web application has been designed for the purpose of collecting user photographs and annotations. A new user is required to register on the website using registration form. A registered user can login into the members area and then create albums, upload photographs to those albums and annotate photographs. Other functionalities in-

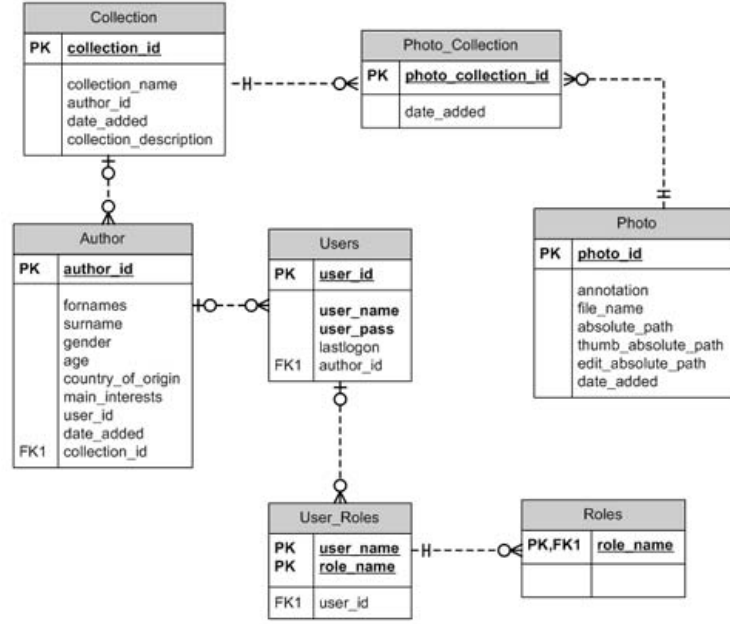


Fig. 4.9: Entity Relationships diagram.

include an album slide show, reading and displaying of EXIF information of a photograph and updating a *UserComment* tag in EXIF header with user annotations. Registration, login and browsing in the members area are all performed on secure connection.

Java technology (Sun Microsystems Inc 2005), specifically java server pages (jsp) and java beans, is used to implement the Collector web application. Figure 4.10 contains a diagram that reflects the structure of the web application and various navigation routes that correspond to user actions. Table .2 in Appendix contains information about the web pages that use java beans and the scope of each java bean.

Postgresql (PostgreSQL Global Development Group 2005) open source database technology is used for data source. Figure 4.9 shows a detailed entity relationships (ER) diagram of the database.

Registration and login are performed over secure connection that uses high-grade encryption of 128 bit and Secure Socket Layer (SSL) technology. Apache server provides a very convenient and simple way to declare pages that must be protected. For this purpose, it is required to specify the `<security-constraint />` xml tag in web.xml file of the Collector web application. Each protected resource should have a corresponding security constraint element. When a user requests a protected resource, the browser presents a user with the certificate that serves as a set of credentials to identify the site. Typically, this certificate must be obtained and signed by the appropriate authority but can also be self-signed. The Collector web application uses self-signed certificate that

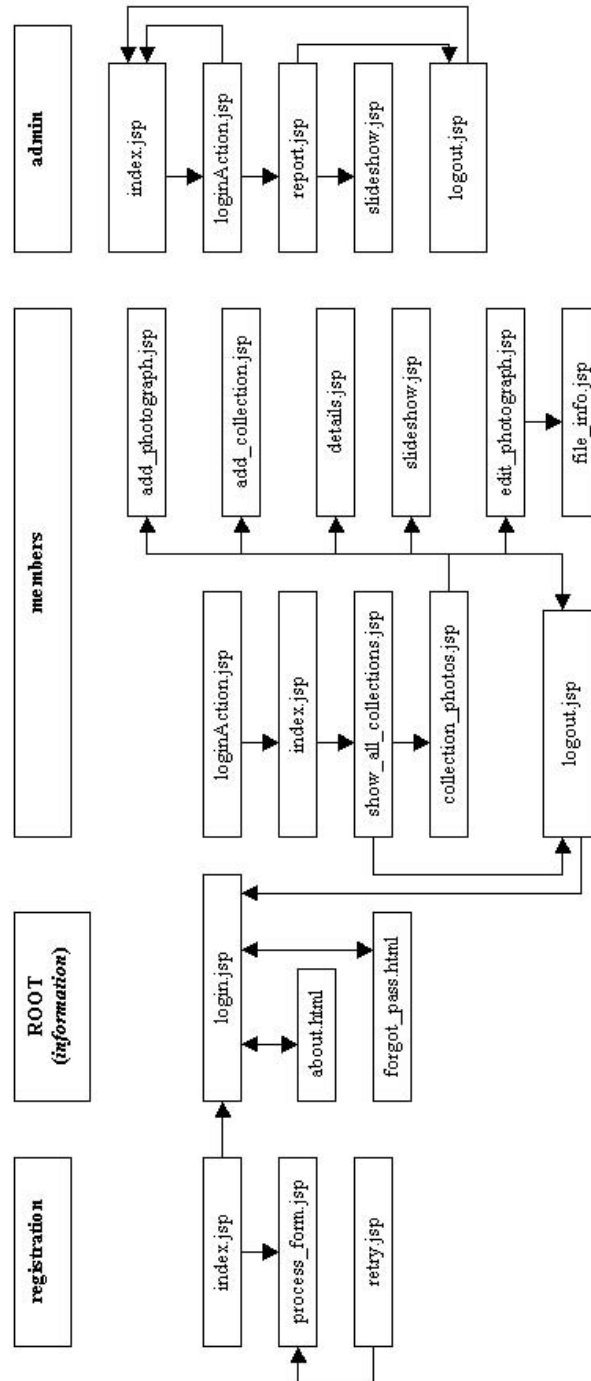


Fig. 4.10: Structure of web pages in the Collector web application.

identifies the organization the website belongs to and the address of this organization.

The Collector web application comprises of four main areas: information area, registration area, members area and administration area. The purpose of each area and its functionalities are outlined below.

The information area consists of three pages: `login.jsp`, `about.html` and `forgot_pass.html`. `login.jsp` is the main page of the web application for both registered users and guests. The url address of the main page for the information area is

<http://stuweb3.cmp.uea.ac.uk/a417556/webapp/>. `login.jsp` provides a short description of the project, confidentiality information, a link to registration page, a link to more information page, and a login form that accepts user email address and password to allow a user to login to the members area.

The registration area consists of three pages: `registration/index.jsp`, `registration/retry.jsp` and `registration/process_form.jsp`. The url address of the main page for the registration area is <http://stuweb3.cmp.uea.ac.uk/a417556/webapp/registration/>.

`registration/index.html` page contains the registration form. The registration form includes required and optional fields. The required fields are email, password, confirm password, age, gender and country of origin. The optional fields include forenames, surname and main interests. `registration/process_form.jsp` page and `registration/retry.jsp` page are not explicitly visible to a user. Their purpose is to validate registration form entries.

A user can enter members area by submitting email address and password on a login form on `login.jsp` page. The login request is processed using `members/loginAction.jsp` (Seshadri 2003) page. The application does not allow a user to use browser page navigation buttons or browser refresh page mechanism to display user information in members area once a user is logged out.

The last login is measured in milliseconds from the 1st of January 1970 and is updated in the database to a current value each time user logs in. The login mechanism allows multiple users to be logged into the same account simultaneously. A user is redirected to `members/index.jsp` page upon successful login. This page contains functionality that identifies the author. Author details are stored in the java bean “Author”. Author id value is stored in another java bean “Manager” that has a session scope and is used for convenient way of extracting various data from the database associated with that author.

The design of the web application separates the user and author. This enables the



Fig. 4.11: Members area navigation menu.

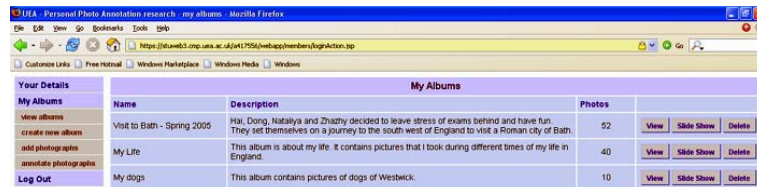


Fig. 4.12: An image screen of members/show_all_collections.jsp web page.

support for various user roles.

The url address of the main page for members area is http://stuweb3.cmp.uea.ac.uk/a417556/webapp/members/show_all_collections.jsp. There are eleven jsp pages in members areas: members/add_collection.jsp, members/add_photograph.jsp, members/collection_photos.jsp, members/loginAction.jsp, members/show_all_collections.jsp, members/edit_photograph.jsp, members/file_info.jsp, members/index.jsp, members/details.jsp, members/logout.jsp and members/slideshow.jsp. In the members area each page displays a menu on the left hand side. Figure 4.11 presents an image of the menu.

The first page displayed to a user is members/show_all_collections.jsp. Figure 4.12 shows an image screen of members/show_all_collections.jsp for a user who has three albums.

On this page a user can delete an album, view album slide show (Arnold 2005), and view photographs of an album. The submenu “annotate photographs” takes a user to members/show_all_collections.jsp?annotationcheck=empty. This page is exactly the same as the members/show_all_collections.jsp page with the only difference that it displays the albums that contain photographs without annotations.

The first thing a new user should do is to create an album. For this task a user must use “create new album” submenu that takes the user to members/add_collection.jsp page. Figure 4.13 contains an image screen of members/add_collection.jsp page.

On this page a user is presented with the new album form. The new album form contains two fields: album name and description. The number of albums that can be created by a user is not limited. The next step is to add photographs. For this task a user must



Fig. 4.13: An image screen of members/add_collection.jsp web page.

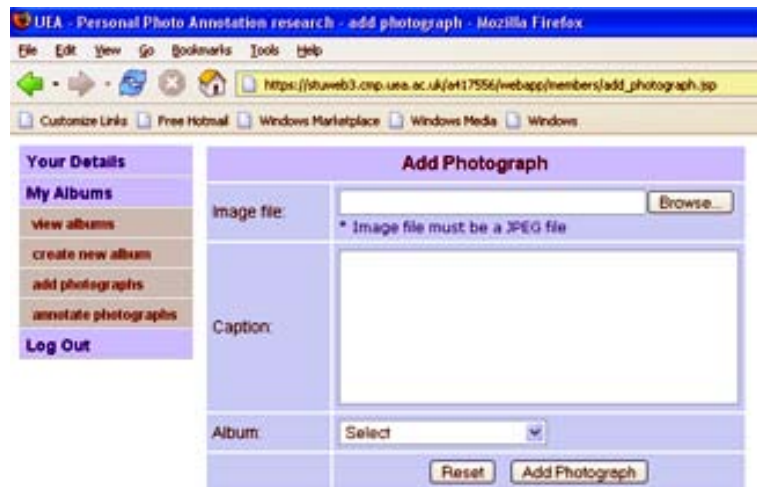


Fig. 4.14: An image screen of members/add_photograph.jsp web page.

choose submenu “add photographs” that takes a user to the members/add_photograph.jsp page. Figure 4.14 contains an image screen of members/add_photograph.jsp page.

members/add_photograph.jsp page consists of add photograph form that contains three fields: image file, caption and album. Only one photograph can be uploaded at a time. Image file must be a JPEG file and can be of any size. The functionality of copying selected by a user image file from a local directory to a remote server is provided by javazoom.upload.UploadBean (JavaZOOM 2005).

Once the image file is copied to the remote directory *UserComment* tag in EXIF header is updated with the values obtained from the “Caption” field on add photograph form. If an image does not contain *UserComment* tag in EXIF header than the web application creates this tag and sets its value to the value obtained from the “Caption” field.

The Collector web application utilizes EXIF class library described in section 4.1 to read and write EXIF header of JPEG files. There is no limit to the number of photograph

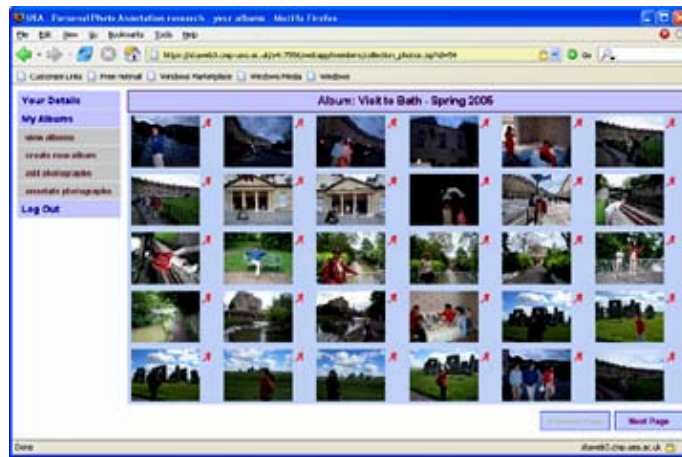


Fig. 4.15: An image screen of members/collection_photos.jsp web page.

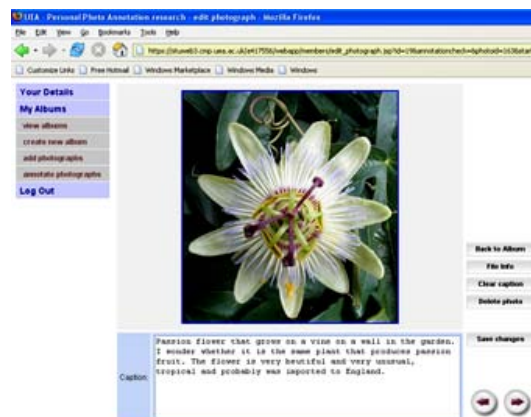


Fig. 4.16: An image screen of members/edit_photograph.jsp web page.

that a user can upload.

A user can view photographs uploaded to a particular album by pressing button “View” on the members/show_all_collections.jsp page. This application takes a user to members/collection_photos.jsp page that displays 30 album thumbnail images per page. Figure 4.15 contains an image screen of an example of this page. Annotated images have letter A next to them.

Clicking an image takes a user to the members/edit_photograph.jsp image page. Figure 4.16 contains an image screen of an example of this page.

At the bottom of the image preview there is a text field that contains current annotation text. On the right hand side of the image preview there is a set of menus: back to album, file info, clear caption, delete photo and save changes. There are also buttons that can be used to navigate to the next and previous photo in the album. “File Info” button displays a page in new window that contains a table filled with EXIF data extracted from the image. Figure 4.17 contains an image screen of an example of this page.

Every time a user modifies an image caption and saves changes *UserComment* tag

IFD Name	Tag Code	Tag Name	Value	Meaning
IFD 0	271	Make	Canon	Canon
IFD 0	272	Model	Canon PowerShot A300	Canon PowerShot A300
IFD 0	274	Orientation	1	The 0th row is at the visual top of the image, and the 0th column is the visual left-hand side.
IFD 0	282	XResolution	180	180
IFD 0	283	YResolution	180	180
IFD 0	296	ResolutionUnit	2	inch
IFD 0	306	DateTime	2005:08:08 20:50:08	2005:08:08 20:50:08
IFD 0	531	YCbCrPositioning	1	centered
IFD 0	34665	ExifFPPointer	185	185
Exif Sub	33434	ExposureTime	1/800	1/800
Exif Sub	33437	FNumber	2.8/5	2.8/5
Exif Sub	36864	ExifVersion	0220	Ver. 2.2
Exif Sub	36867	DateTimeOriginal	2003:12:07 14:55:17	2003:12:07 14:55:17
Exif Sub	36868	DateTimeDigitized	2003:12:07 14:55:17	2003:12:07 14:55:17
Exif Sub	37121	ComponentsConfiguration	1 2 3 0	YCbCr
Exif Sub	37122	CompressedBitsPerPixel	3/1	3/1
Exif Sub	37377	ShutterSpeedValue	309/32	309/32
Exif Sub	37378	ApertureValue	159/32	159/32
Exif Sub	37380	ExposureBiasValue	0/1	0/1
Exif Sub	37381	MaxApertureValue	59/16	59/16

Fig. 4.17: An image screen of `members/file_info.jsp` web page.

Your Details	
Forenames	Natalya
Surname	Alexander
Gender	female
Age	26 - 35
Country of Origin	Ukraine
Main Interests	Reading, Music, Dancing, Walking, Swimming, Learning new things, All
User Name	n.alexander@ues.ac.uk

Fig. 4.18: An image screen of `members/details.jsp` web page.

in EXIF header is updated. A user can also click on the preview image to open a new window with the original image. The *UserComment* tag in EXIF header is updated only in the original image.

The image files are stored in directories. The path template to an uploaded image is `/members/user_images/author_id/album_id/photo_id/photo.jpg`, where text in italic font is replaced with the actual values. Two extra images created in addition to the original image. The first image is the thumbnail image with prefix “thumb_” used on the `members/collection_photos.jsp` page as a preview image. The second image is a reduced in size image with prefix “edit_” used on the `members/edit_photograph.jsp` page. The maximum width and height of this image does not exceed 512 pixels. `createThumbnail` method in `PhotoUpload` java bean is used to reduce the size of an image (dmitri_don 2002).

In the members area a user can also see the details supplied during registration by selecting “My Details”. Figure 4.18 contains an image screen of the `members/details.jsp` page.



Fig. 4.19: An image screen of `admin/report.jsp` web page.

The last part of the Collector web application is the administration area. The url address of the main page of the administration area is

<http://stuweb3.cmp.uea.ac.uk/a417556/webapp/admin/>.

There are five jsp pages in this area: `admin/report.jsp`, `admin/index.jsp`, `admin/slideshow.jsp`, `admin/logout.jsp` and `admin/loginAction.jsp`. On `admin/index.jsp` page there is a login form for an administrator. If authentication is successful than the administrator is redirected to `admin/report.jsp`. On this page an administrator can view user details, albums, photographs and annotations. One user per page is displayed. Figure 4.19 shows an image screen of an example of the `admin/report.jsp` page.

The accompanying CD contains source files for java beans, java server pages files, css file, jar library files, sql script file for re-creating tables in postgresql database, the csv files with the data from the database tables, and `user_images` directory with images supplied by users.

5. EXPERIMENT

The aim of the experiment is to obtain the maximum amount of data. All photographs and annotations were obtained through the Collector web application. The main method used for getting people to join, upload and annotate their photographs was emailing. Several identical emails were sent to different groups of people. Altogether 120 people were contacted. The desired amount of photographs per user was set to 20 and specified in the email. The addresses of the recipients comprised mainly of family, friends and classmates. Other methods of data acquisition included telephone calls and personal conversations. It took almost three weeks to obtain the amount of data listed in Table 5.1. Some users required a follow-up call or a reminder while others submitted photographs within the next 3 days of the email.

Entity	Count
Users	27
Albums	49
empty albums	3
avg. albums per user	1.8
Photos	603
with captions	566
without captions	37

Tab. 5.1: Statistics.

The email text contained some brief information about the project, the address of the website, instructions on how to use the website and the preferred number of photographs and annotations. It also contained security and confidentiality information. One user sought confirmation in person that the submitted photographs will not be published anywhere without permission from friends who also appeared on the photographs. There were several people who required around 40 minutes of additional explanation of the project but when joined uploaded no more than three photographs or no photographs at all. Overall, 27 people have joined the website which is 22.5%, 27 joins out of potential 120. This ratio could have been significantly improved if some kind of reward was offered. Another possible explanation for such a low ratio is the fact that the requested

photographs are personal photographs and some people may be reluctant to share them. Three persons claimed that they had only a few personal digital photographs. One person also complained that it is very difficult to think of any annotations, particularly when English is not a native language. Table 5.1 contains primary statistical information of the volume of entries and how complete they were.

The most valuable information about the users which is collected upon their registration consists of **age**, **country of origin**, **gender** and **main interests**. For age a user is required to select an age group rather than an exact number of years. Table .4 in the Appendix contains a list of age groups available for selection in the “Age” field on the registration form. Table .3 in the Appendix contains a list of names of countries available for selection in the “Country of Origin” field on the registration form.

6. ANALYSIS

6.1 *Data familiarisation, attribute enrichment and cleansing*

The initial stage of the analysis consists of data familiarisation. The collected photographs and corresponding annotations are closely studied. Following the studying of the records, a number of additional attributes for description of photographs and annotations are proposed. The values of attributes are hand labelled for each record.

The first proposed attribute is the *Structure* with the values of **who**, **what**, **location**, **event**, **action**, **timeline** and **emotion**. The *Structure* attribute can take multiple values. The *Structure* is present in all annotations within the dataset. It is a generalised view of the contents of an annotation. The observations made during the data familiarisation stage propose that the *Structure* attribute can be used to describe the contents of any annotation in the dataset. However, the set of values of the *Structure* attribute is not suitable to generalise the contents of a long story without some mechanism of structure parsing. The *Structure* attribute is useful for analysis because it provides a common ground for comparing the contents of annotations.

The second proposed attribute is *Artistic*. This attribute can take only boolean values, true or false, and denotes whether the annotation text is artistic or not. There are annotations that simply state who is featured in the image, when it was taken, where it was taken, and what event it signifies. These annotations are mere descriptors. Artistic annotations are creative and often humorous. All artistic annotations contain emotions. However, annotations that contain emotions are not always artistic. Artistic annotations are not just references to past times, people, animate and inanimate objects, but also a tool to involve the viewer's senses.

The third proposed attribute is the *Length*. This attribute can take only one of the following values: **short**, **medium** or **long**. Short annotations consist of no more than 5 words. Medium annotations contain between 6 and 15 words and typically consist of no more than two sentences. Long annotations contain 16 words or more. Long annotations

often contain a story behind the photograph and take more effort to complete.

It is also useful to obtain part-of-speech tags for each annotation. Part-of-speech tags are used to determine the proportion of word classes for annotations and the distribution of a particular word class in the *Artistic*, *Length* and *Structure* attributes. For this purpose we use the transformation-based or “Brill” part-of-speech tagger (Brill 1995) for Windows (Ghadirian 2004) and Penn tree bank tagset (Mitchell et al. 1993) and its most important tags (mozart-oz.org 2004). The result is a string of tags for each annotation separated by space character. The order of tags in the string is the same as the order of words in the corresponding annotation. The accuracy of the tags is checked for each annotation and only a few corrections are made. Further processing includes splitting the string of tags into separate tags to produce a count of each tag for each annotation.

Another useful information can also be extracted by analysing the contents of the images and labelling them with the words that describe major objects present in the images. Some examples of such words are cars, trees, leaves, people, buildings, river and sky. This information provides an insight into the types of personal photographs people take. For this task all records are hand labelled.

It is also interesting to find out how much information there is in each annotation. For this task we use entropy from Claude Shannon Information Theory as a measure of information. Entropy is used to calculate how much information is carried by each word (Belew 2000). It is calculated using the following three equations. Equation 6.1 is used to calculate the amount of **noise** in bits for a particular word. Equation 6.2 is used to calculate the amount of **signal** in bits for a particular word. The signal is then used to calculate the signal weight of a particular word in an annotation. The equation 6.3 is used for this calculation.

$$\langle Noise_k \rangle = \langle (p_k \log(1/p_k)) \rangle = \sum_d \frac{f_{kd}}{f_k} \log \frac{f_k}{f_{kd}} \quad (6.1)$$

$$Signal_k = \log f_k - Noise_k \quad (6.2)$$

$$w_{kd} = f_{kd} * Signal_k \quad (6.3)$$

where f_k is a number of times word k appears in all annotations, f_{kd} is a number of times a word appears in a particular annotation.

The sum of signal weights of annotation words normalised by the number of words in the annotation is used as a measure of information conveyed by the annotation.

During the cleansing stage records are checked for empty annotation fields and duplicate records. Records that contain no annotations are removed. Duplicate records are also removed. The resulting dataset contains 559 records.

The last step is the preparation of the .csv file used as a data source for clustering in Clementine knowledge-discovery in databases (KDD) environment. A series of cascading queries are applied to obtain necessary data in a suitable for importing to a .csv file format. There are 23 fields in total. However, not all fields are going to be used for clustering. The reason for including 23 fields is that Clementine has powerful tools for data visualisation which makes it convenient for producing graphs for all fields of interest. Table .1 in Appendix contains data dictionary of the fields imported to the .csv file. Note, that the asterisk symbol next to the name of a field indicates that this field is used in clustering.

6.2 Data Visualisation

In this stage several attributes are analysed using various graphs and charts. The analysis begins with word classes. There are four main classes of words in the English language: **nouns**, **verbs**, **adjectives** and **adverbs**. Nouns can also be categorised into proper nouns and common nouns. For the analysis of word classes, all word classes are grouped into 6 categories: **proper nouns**, **common nouns**, **verbs**, **adjectives**, **adverbs** and **other**. The total number of instances of each category can be found in Table 6.1 below. Figure 6.1 contains a pie chart of word classes based on values in Table 6.1. It is also useful to find the proportion of just the four main word classes of English language. Figure 6.2 contains a pie chart of the four main word classes of English language but nouns are split into common and proper nouns.

According to Figure 6.1 38% of all word classes belongs to category **other**. Category **other** contains word classes that are not useful for describing entities and their characteristics in annotations.

Word Class	Count
Proper nouns	627
Common nouns	1036
Verbs	478
Adjectives	282
Adverbs	127
Other	1524

Tab. 6.1: Word classes with count.

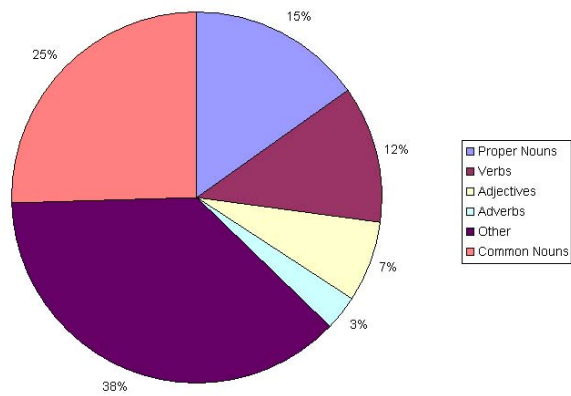


Fig. 6.1: Pie chart of word classes.

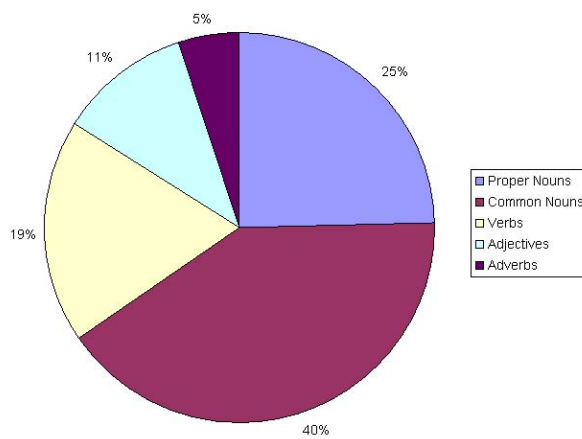


Fig. 6.2: Pie chart of the four main word classes of the English language.

The largest word class category is nouns. Combined common and proper nouns comprise 40% of all word classes found in annotations. It is also interesting to observe that 15% of all word classes belongs to proper nouns. Proper nouns also comprise 37% of all nouns, 627 proper nouns of total 1663 nouns. Moreover, the percentage of proper nouns is more than the percentage of verbs, adverbs or adjectives. According to Figure 6.2 proper nouns comprise 25% of the four main classes of the English language and are the second largest group.

Verbs are the second the second largest category of the four main word classes. Verbs comprise 12% of all word classes and 19% of the four main word classes. Adjectives are the third largest category of the four main word classes. Adjectives comprise 7% of all word classes and 11% of the four main word classes. The last and the smallest word class category of the four main word classes is adverbs. Adverbs comprise only 3% of all word classes and 5% of the four main word classes.

Nouns is the most frequently occurring word class. Nouns are very important in annotations. Furthermore, proper nouns comprise just over a third of all nouns. This suggests that the presence of names of entities such as people and place is very important in annotations. Verbs and adjectives are the two other important groups of word classes. The least significant group of the four main word classes in annotations is adverbs.

As mentioned in Section 5 attributes **age**, **country of origin** and **gender** are very important for understanding how various groups of people annotate their photographs. The next stage of data visualisation contains analysis of the distribution of annotations for each of these attributes.

Figure 6.3 shows the distribution of annotations for age groups. According to Figure 6.3, there are two major age groups: 19-25 and 26-35. 88.55% of all annotations is supplied by the representatives of these groups, where 34.7% belongs to age group 19-25 and 53.85% belongs to age group 26-35. The representatives of age groups under 18-24, 36-45, 46-55, 56-65, 65-75, 75 and over, supplied only 11.45% of all annotations. There are no annotations supplied by representative of age groups 56-65, 65-75, 75 and over. Therefore, the eldest users of the Collector web application fall into the 46-55 age group. The conclusion is that there are not enough annotations from representatives of all age groups. Thus, age attribute is not used in clustering or any further analysis.

Figure 6.4 shows the distribution of annotations in countries. Similarly to age groups,

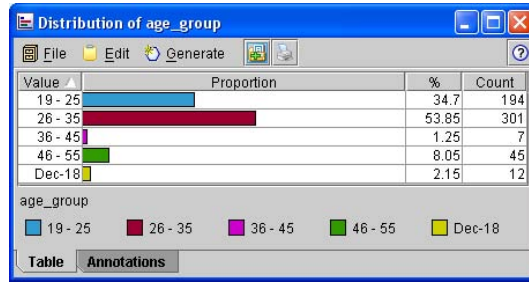


Fig. 6.3: Distribution of annotations in age groups.

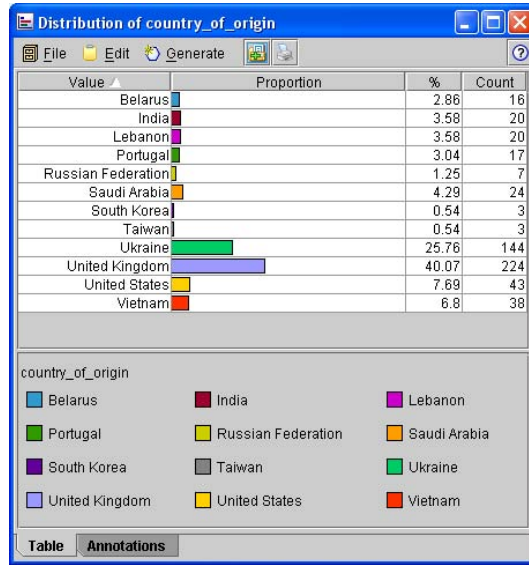


Fig. 6.4: Distribution of annotations in Countries of Origin.

there are two countries that comprise the majority of 65.83%, which is a sum of percentages of annotations supplied by users originated in Ukraine and United Kingdom. The remaining 12 countries are allocated only 34.17% of all annotations. Furthermore, there are tens of countries for which there are no annotations. This leads to the conclusion that there are not enough annotations from the variety of countries, such as from countries with occidental and oriental cultures. Thus, the country of origin attribute is not going to be used further in the analysis and clustering.

Figure 6.5 shows the distribution of annotations for gender. In this figure, the distribution of annotations for males and females is close to equal, 54.38% of females and 45.62% of males. This means that there is a sufficient percentage of annotations from both females and males to use in the analysis and clustering.

The next selected attribute for analysis is the *Structure*. Figure 6.6 contains a pie chart of the *Structure* attribute values found in annotations. In addition to Figure 6.6 there is a Table .5 in Appendix that contains a list of distinct values of the *Structure* attribute with count for each value.

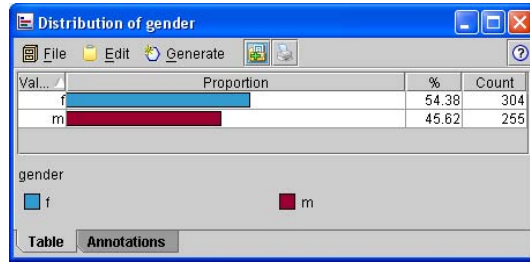


Fig. 6.5: Distribution of annotations in gender.

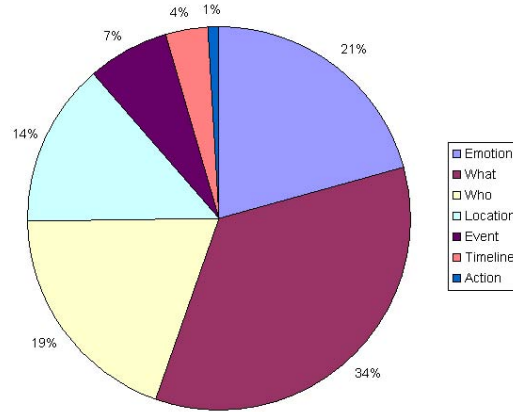


Fig. 6.6: Pie chart of the *Structure* attribute values.

According to Figure 6.6 there are four values of the *Structure* attribute that occur the most frequently: **what**, **who**, **emotion** and **location**. The value **what** occurs the most frequently and comprises 34% of all values. The next most often occurred value is **emotion** at 21%. After the **emotion** there is the value **who** at 19% and then the value **location** at 14%. The value **event** is only present in 7% of annotations. The percentage of the values **action** and **timeline** is very small and combined equals to 3%. Interestingly, the percentage of the value **emotion** is more than the percentage of the value **who** by 3%. This could be contributed by the value **emotion** being present in annotations that contain the values **who** and **what**. Table .5 in Appendix shows that there are four main groups of combinations of values of the *Structure* attribute. These groups are **what**, **emotion** and **what**, **who**, and **emotion**. It is interesting to observe that in these four groups the value **emotion** occurs either with the value **what** or on its own but not with the value **who**. Further investigation is conducted to find how accurate this observation is for all combinations of values. The combination of the values **emotion** and **who**, and **emotion** and **what** was counted in all records. The combination of the values **emotion** and **who** occurs 54 times. The combinations of the values **emotion** and **what** occurs 81 times, which is 50% more than the combination of the values **emotion** and **who**. Due to the size of the experimental dataset, the accuracy and the truthfulness of this observation cannot be confirmed but

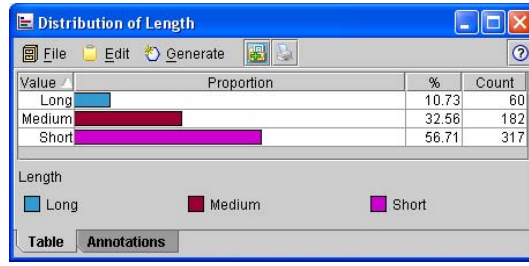


Fig. 6.7: Distribution of annotations in the *Length* attribute.

provides an interesting suggestion for future investigations.

The next analysed attribute is the *Length* attribute. Figure 6.7 contains distribution of annotations in three values of *Length* attribute: **short**, **medium** and **long**.

According to Figure 6.7 56.71% of all annotations contain no more than 5 words. The percentage of medium length annotations is 32.56%. These annotations contain between 6 and 15 words. Only 10.73% of all annotations contains more than 15 words. The conclusion is that short and medium annotations cover almost 90% of all annotations and short annotations cover just over a half of all annotations.

Objects and people that appear on the images provide some insight into the types of photographs people. All photographs are hand labelled with the names of main objects. There are 144 distinct objects in total. This is a very detailed information. The objects are further generalised into the following categories: clothes, other, man made objects, food, works of art, transport objects, water, constructed outdoor objects, landscape, animals, household objects, constructed indoor objects, sky, buildings, vegetation, and people. The original list of objects with the corresponding generalised categories of objects can be found on the accompanied CD. Table .6 in Appendix contains a list of the generalised categories of objects with count. According to Table .6 there are two major categories of objects: people and vegetation. This is also reflected in the pie chart in Figure 6.8.

Annotations that contain value **emotion** in the *Structure* attribute are subdivided further into artistic and not artistic. Figure 6.9 contains the distribution graph of annotations in the *Artistic* attribute. It is also interesting to find in Figure 6.10 that only 47.85% of annotations that contain emotions are artistic.

Another important attribute is *Information*. Figure 6.11 contains a histogram of *Information* attribute measured in bits with the *Length* attribute selected as colour overlay. There are two observations based on this histogram.

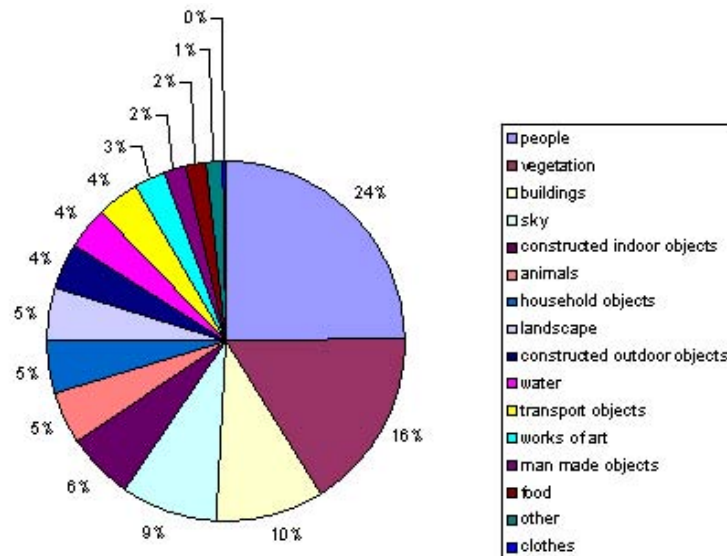


Fig. 6.8: Pie chart of generalised categories of objects found in photographs.

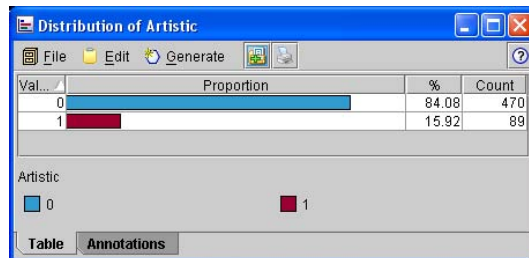


Fig. 6.9: Distribution of annotations in the *Artistic* attribute.

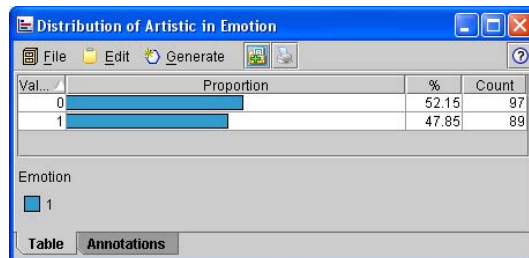


Fig. 6.10: Distribution of artistic annotations in the value *emotion* of the *Structure* attribute.

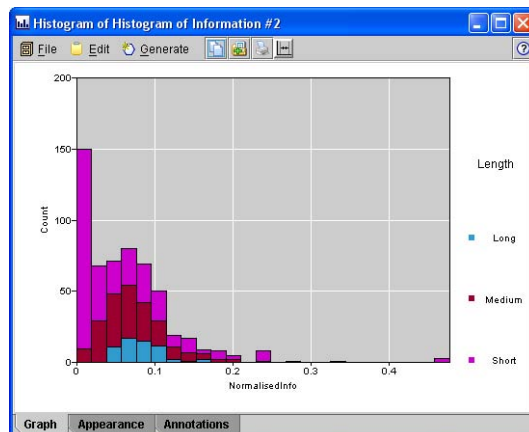


Fig. 6.11: Histogram of *Information* attribute values measured in bits with the *Length* attribute selected as colour overlay.

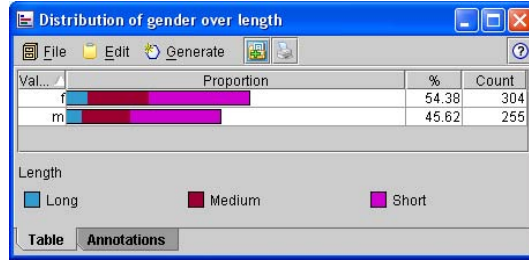


Fig. 6.12: Distribution of annotations in gender with the *Length* attribute selected as colour overlay.

The first observation is that more than 80% of annotations are contained within the range of 0 and 0.1 bits with the majority of short annotations. The second observation is that the majority of annotations between 0 and 0.05 bits are short annotations, while the majority of annotations between 0.05 and 0.1 are medium and long annotations. However, based on visual analysis of the distribution of the *Length* attribute in the values of *Information* attribute, there is no significant variation in the percentages of the *Length* attribute values between bins. The only exception is for bin between 0.2 to 0.3 which contains mainly short annotations.

Based on the analysis in this section, gender is the attribute that is used for further analysis and clustering. From gender it is possible to find out how females and males annotate their photographs, what are the differences and similarities for these groups of people and if there are any subgroups within these groups. In the data visualisation stage, it is useful and interesting to analyse the distribution of annotations of males and females in the following attributes: the four main word classes of English language, the value **emotion** of the *Structure* attribute, the *Length* attribute and the *Artistic* attribute. Distribution graphs are used for this task.

The first distribution graph shown in Figure 6.12 contains the distribution of annotations in gender with the *Length* attribute selected as colour overlay. According to Figure 6.12, there are no interesting patterns because the distribution of length appears to be almost equal for annotations that belong to both males and females.

Figure 6.13 contains the distribution of annotations in gender with the *Artistic* attribute selected as colour overlay. According to this figure, artistic quality is almost equally distributed between males and females and there are no interesting patterns.

Figure 6.14 contains a distribution of annotations in gender with the value **emotion** of the *Structure* attribute selected as colour overlay. According Figure 6.14, annotations

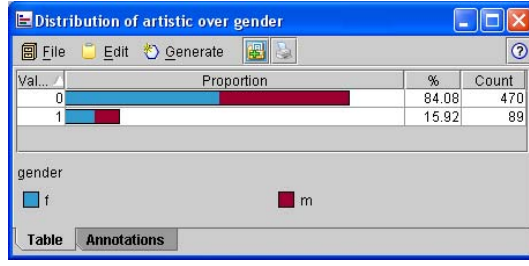


Fig. 6.13: Distribution of annotations in gender with the *Artistic* attribute selected as colour overlay.

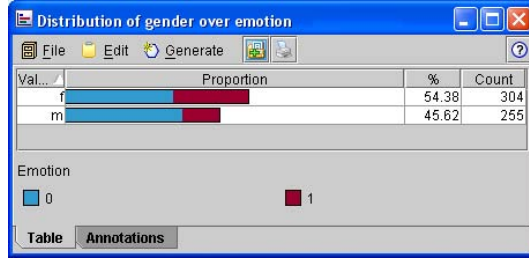


Fig. 6.14: Distribution of annotations in gender with the value **emotion** of the *Structure* attribute selected as colour overlay.

supplied by females contain more annotations with the value **emotion** of the *Structure* attribute than annotations of males.

This is also reflected in Table 6.2 where the percentage of annotations by females that contain the value **emotion** is 41.11% and the percentage of annotations of males that contain the value **emotion** is 23.92%.

This observation is related to the observation found in pie charts of the main four word classes for annotations by females and males in Figures 6.15 and 6.16. According to the pie charts in Figures 6.15 and 6.16, male annotations contain 10% more nouns than female annotations, 3% less verbs, 5% less adjective and 2% less adverbs. The reduction in the use of adjectives means reduction in description of qualitative and quantative characteristics of entities in annotations. The increase in use of nouns means that there are more references to entities.

Gender	Total Count	With Emotions	%age of With Emotions
males	255	61	23.92%
females	304	125	41.11%

Tab. 6.2: The percentage and count of annotations that contain the value **emotion** of the *Structure* attribute in annotations by females and males.

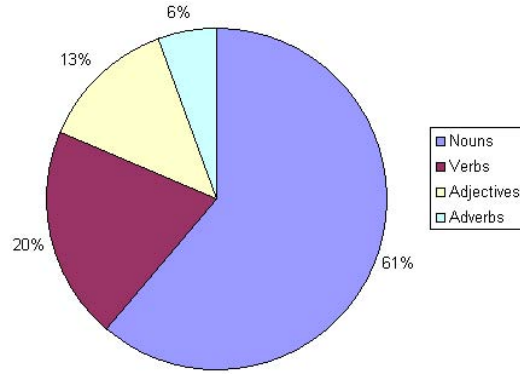


Fig. 6.15: Pie chart of the four main word classes of the English language in annotations by females.

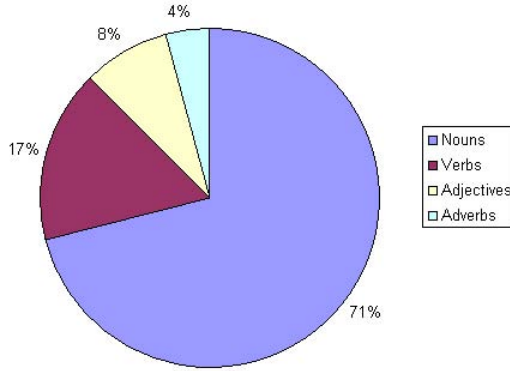


Fig. 6.16: Pie chart of the four main word classes of the English language in annotations of males.

6.3 Clustering

Clustering is used to group records based on their similarity. Records within a group are similar to each other but are different from records of another group (Han and Kamber 2001).

For the clustering task K-Means algorithm in the Clementine KDD environment is used. In the data dictionary provided in Table .1 in Appendix asterisk symbol next to the name of the attribute indicates that this attribute is used in clustering. For K-Means algorithm it is required to specify the number of clusters. The task is to find the number of clusters that are of high quality in term of inter and intra cluster similarity and produce some interesting patterns. Clementine provides intra and inter proximity values for each cluster. This information is useful for evaluation of the quality of clusters.

Initial number of clusters is set to 4. Then, the number is changed to 3 and 5. There are three models of clusters in total: 3 clusters model, 4 clusters model and 5 clusters model. The next stage consists of determining the best quality model. For this task we use a set of figures that visualise the distances for intra and inter similarity for each

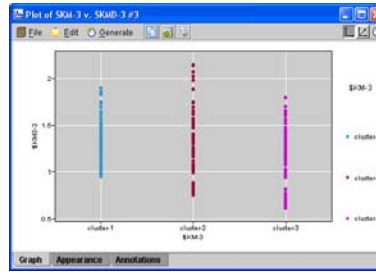


Fig. 6.17: Plot of proximity values of the model with 3 clusters for intra cluster analysis.

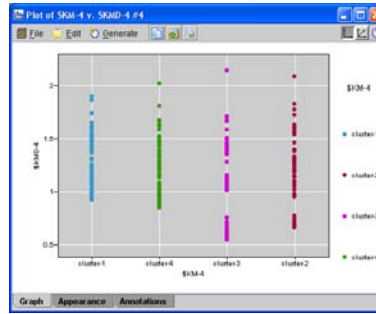


Fig. 6.18: Plot of proximity values of the model with 4 clusters for intra cluster analysis.

model. In addition to the figures, there is also Table .7 in Appendix that contains the values of distances between cluster centroids for each cluster.

Figures 6.17, 6.18, 6.19 shows a plot of proximity values for intra cluster similarity. The best model for intra similarity is the model with 3 clusters. The next best model is the model with 4 clusters.

Figures 6.20, 6.21, 6.22 shows the plot of proximity values for inter cluster similarity. Based on visual analysis the best two models are 3 cluster model and 4 cluster model.

For both models there are two clusters that are very close to each other. However, according to the information in Table .7, the distance measured at 0.15604 between the two closest to each other clusters in 3 cluster model is 0.013854 less than the distance measured at 0.170058 between the two closest to each other clusters in 4 cluster model. The model with 4 clusters is selected based on the results of inter and intra similarity

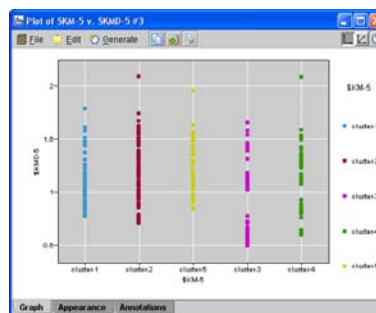


Fig. 6.19: Plot of proximity values of the model with 3 clusters for intra cluster analysis.

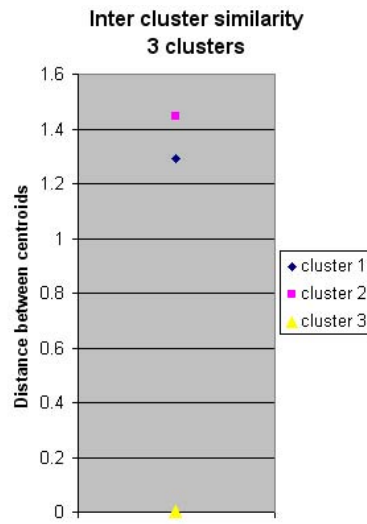


Fig. 6.20: Plot of proximity values of the model with 3 clusters for inter cluster analysis.

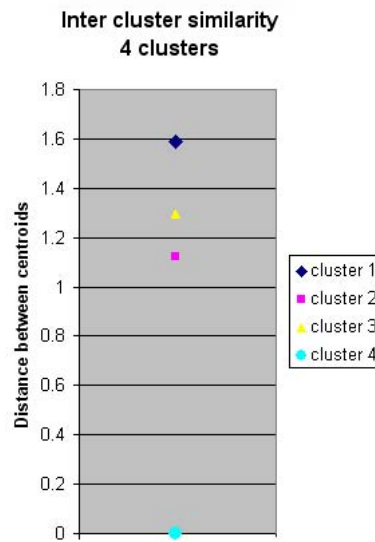


Fig. 6.21: Plot of proximity values of the model with 4 clusters for inter cluster analysis.

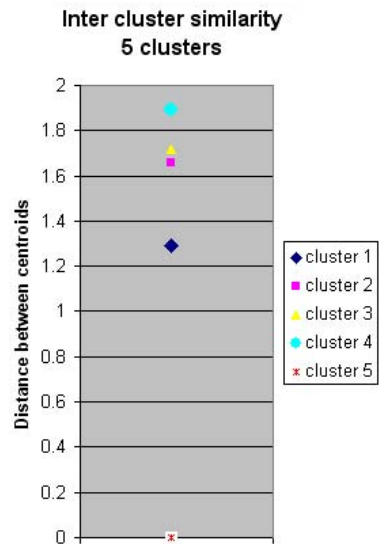


Fig. 6.22: Plot of proximity values of the model with 3 clusters for inter cluster analysis.

Gender	%age in Cluster 1	%age in Cluster 2	%age in Cluster 3	%age in Cluster 4
females	65%	82%	48%	43%
males	35%	18%	52%	57%

Tab. 6.3: Percentage of females and males in each cluster.

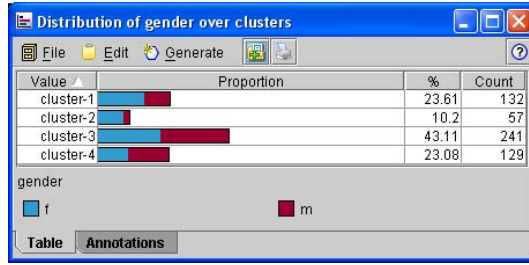


Fig. 6.23: Distribution of clusters with gender selected as colour overlay.

analysis.

6.4 Analysis of Clusters

The first analysed attribute is gender. Figure 6.23 contains the distribution graph of clusters with gender selected as colour overlay. In supplement to Figure 6.23 there is also Table 6.3 where the percentage of each gender per cluster is shown.

According to Table 6.3, in cluster 3 there is 4% more annotations by males than annotations by females. In this cluster the percentage of annotations by females and males is almost equal. In cluster 4 there is 15% more annotations by males than annotations by females. In this cluster annotations by males represent the majority of annotations. In cluster 1 there is 30% more annotations by females than annotations by males. This is a considerable difference. Lastly, in cluster 2 there is 65% more annotations by females than annotations by males. This cluster mainly consists of annotations by females.

The next analysed attribute is the value **emotion** in the *Structure* attribute, the *Artistic* attribute, the *Length* attribute, the main four word classes of the English language and the *Structure* attribute.

Figure 6.24 contains the distribution graph of clusters with the value **emotion** of *Structure* attribute selected as colour overlay.

In Figure 6.24 four clusters can be combined into two groups. The first group of clusters consists of clusters 3 and 4. In this group, annotations contain no emotions. The second group of clusters consists of clusters 1 and 2. In cluster 1 all annotations contain emotions. 49 annotations of 57 annotations in cluster 2 contain emotions, which is 86% of

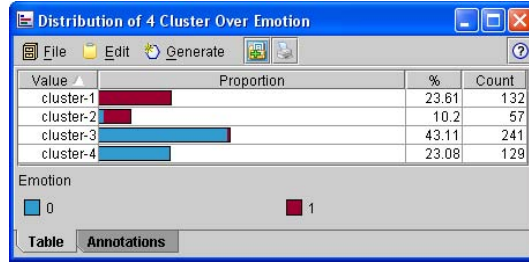


Fig. 6.24: Distribution graph of clusters with the value *emotion* of the *Structure* attribute selected as colour overlay.

annotations with emotions. These two clusters mainly consist of annotations by females, particularly cluster 2. This observation suggests that there are more annotations with emotions in annotations by females than in annotations by males. These two clusters do not entirely consist of annotations by females. However, these clusters include all annotations with emotions. According to Table 6.2 there are 61 annotations by males that contain emotions and 125 annotations by females that contain emotions, which is 186 annotations that contain emotions in total. Therefore, the percentage of annotations by males in all annotations with emotions is 32.8%.

Figure 6.25 contains the distribution graph of clusters with artistic attribute selected as colour overlay. In this figure, cluster 1 contains 95.5% of all artistic annotations, which is 85 artistic annotations of 89 artistic annotations in total. Cluster 2 only 4.5% of all artistic annotations, which is 4 artistic annotations of 89 artistic annotations in total. However, 86% of annotations in cluster 2 contains emotions, which is 49 annotations with emotions out of 57 annotations in total in cluster 2. This is a considerable amount of annotations with emotions. This cluster has a group of annotation that are not artistic but only emotional. According to Table 6.3 82% of annotations in this cluster belong to females. A very different result is in cluster 1 where all annotations are artistic. In this cluster 65% of annotations belongs to females and 35% of annotations belongs to males. This suggests that there is a group of annotations that consists of females and males and contains artistic annotations but the majority of annotations are annotations by females. Cluster 3 and cluster 4 do not have any artistic annotations in them. In these clusters the proportion of annotations by females and males is almost equal. These groups of annotations do not have any emotions and belong both to females and males with equal distribution in each group.

The next analysed attribute is the *Structure* attribute. Figure 6.26 contains four pie

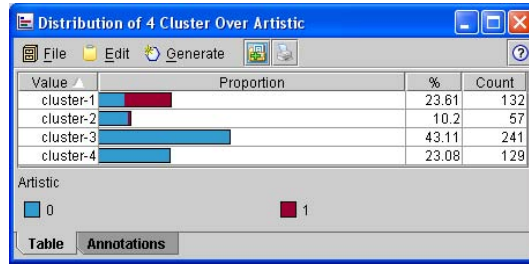


Fig. 6.25: Distribution graph of clusters with artistic attribute as colour overlay.

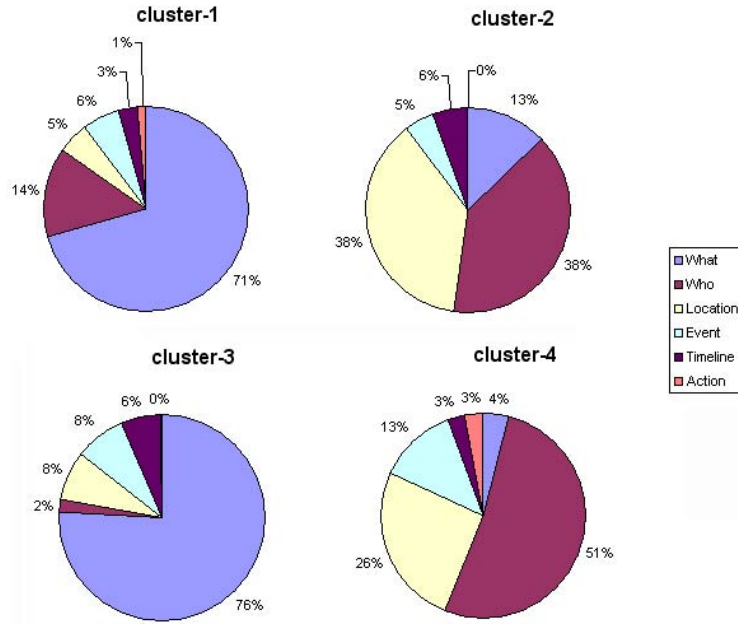


Fig. 6.26: A set of pie charts for each cluster with the percentages of the values of the *Structure* attribute in each cluster.

charts that show the percentage of each value of the *Structure* attribute for every cluster. According to this figure, clusters 1 and 3 are similar because both clusters contain over 70% of the value **what**. In cluster 1 almost all annotations are emotional and artistic while cluster 3 contains no artistic or emotional annotations. In section 6.2 we have managed to relate the value **what** of the *Structure* attribute to emotional annotations. Furthermore, cluster 1 contains artistic annotations and in this cluster the percentage of the value **what** is 71%. In Cluster 3 76% of annotations contain mainly the value **what** but do not convey any emotions. Cluster 2 and cluster 4 are similar to each other because in both clusters over 76% of annotations contain the values **who** and **location**. Moreover, the percentage of the values **who** and **location** in cluster 2 is equal.

It is also interesting to find out what word classes are present in each cluster and the percentage of each word class in a cluster. Figure 6.27 contains four pie charts that show the percentage of the four main word classes in each cluster.

According to this figure the percentage of common nouns does not significantly differ

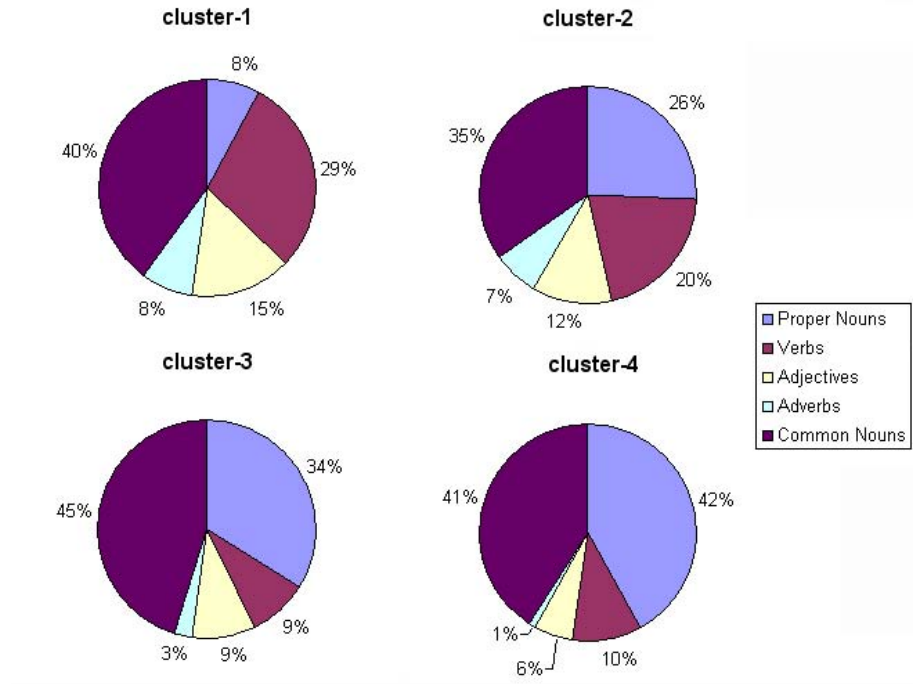


Fig. 6.27: A set of pie charts for each cluster with the percentages of the four main word classes of the English language in each cluster.

between clusters and stays within the range of 35%-45%. The percentage of proper nouns in each cluster grows considerably from cluster 1 to cluster 4. A jump in the percentage of proper nouns is observed starting from cluster 2. It is only 8% in cluster 1, which is the only cluster with artistic and emotional annotations. In cluster 2 it is 26%, which is 14% more than in cluster 1. This cluster contains mainly emotional annotations but only 4.5% of all artistic annotations. In cluster 3, the percentage of proper nouns is 34%, which is 26% more than in cluster 1 and 8% more than in cluster 2. The percentage of proper nouns in cluster 4 is 42%, which is 34% more than in cluster 1, 16% more than in cluster 2 and 8% more than in cluster 3. Clusters 3 and 4 have no emotional or artistic annotations. On the contrary, the number of verbs and adjectives increases with the increase of the number of emotional and artistic annotations. Another observation is that the fewer proper nouns there are in a cluster the more artistic and emotional the annotations are in that cluster.

A further insight into the annotations is provided in Table .8 in Appendix. This table contains the percentages of generalised categories of objects appeared on the photographs for each cluster. From the analysis above we have established that cluster 1 and cluster 2 contain emotional annotations and cluster 1 also contains artistic annotations. The first interesting observation based on information in Table .8 is that cluster 1 contains the largest proportion of animals in comparison to the remaining three clusters. This

means that perhaps artistic annotations are humorous and relate to animals. The second observation is that cluster 4 contains the largest proportion of people in comparison to the remaining three clusters. This cluster has no emotional or artistic annotations.

In conclusion, we have discovered four groups of annotations. The summarised characteristics of each group are outlined below.

The first group of annotations consists mainly of annotations by females that are artistic and emotional, long in length, with a lot common nouns, verbs and adjectives, but hardly any proper nouns. These annotations are mainly used to describe what and not who. There are a lot of representatives of animals and vegetation. The second group of annotations consists of annotations of both females and males with the majority of annotations by females. These annotations convey emotions but are not artistic. They mainly describe who and location, and use a lot of proper nouns, common nouns, verbs and adjectives. Main objects described by these annotations are people, buildings, vegetation and sky. This suggests that the photographs that belong to this group of annotations are taken in the urban environment. The third group of annotations is almost equally split between annotations by females and males. The annotations in this group contain the largest percentage of proper nouns in comparison to other groups. This group also contains the largest majority of nouns in comparison to other groups. Besides, the percentage of proper and common nouns is almost equal. The nouns in these annotations mainly refer to animate and inanimate objects and not people. The annotations in this group do not contain emotions, are not artistic and are mainly of short and of medium length. The fourth and the last group of annotations consists of annotations by males and females with the majority of annotations by males. These annotations are mainly short and the objects that they describe are mostly people and vegetation. They convey no emotions and are not artistic. The majority of these annotations contains descriptions of who is on the photograph and the location the photograph was taken at. Nouns comprise almost 90% of all word classes with equal distribution between common and proper nouns.

7. PRINCIPLE COMPONENT ANALYSIS (PCA)

Principle Component Analysis is a statistical technique that is used to find patterns in data. PCA is looking for the underlying factors that describe a number of dimensions. The number of dimensions is reduced by replacing the dimensions that have an underlying factor, or a principle component, with the value of this factor.

For PCA analysis it is required to calculate the covariance matrix for dimensions and then calculate eigenvalues and eigenvectors for this matrix. The eigenvectors with the highest eigenvalues are the principle components and represent the most significant relationships in data. Data is now expressed in terms of eigenvectors and each eigenvector is an axis. The elements of an eigenvector are the values for each dimension. The higher the value of the element in an eigenvector the stronger the link of that element with the axes the eigenvector represents (Smith 2002).

In this project PCA is used to discover any underlying factors for the four main word classes of the English language and word count. Note that for PCA analysis the percentage of each word class in an annotation is used and the percentage for nouns is split into the percentages of proper and common nouns. Clementine KDD provides a PCA/Factor used for this task. There are six principle components discovered. Table 7 provides information about the percentage of variance captured by a particular component and cumulative percentage of variance for components. According to Table 7 components 1, 2 and 3 capture 70% of the total variability. Moreover, eigenvalues for components are the highest of all 6 components. Kaiser's Criterion (Field 2004) is one of the measures for selecting significant principle components. It suggests the selection of components with eigenvalues that are more than 1. There are only 3 components that satisfy this criterion: component 1, 2 and 3. High eigenvalues and high coverage of the total variability of components 1 and 2 indicate that these components represent the most significant relationships between data.

Table 7 contains eigenvectors for the first five components. For the first component

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	1.621	27.019	27.019
2	1.515	25.248	52.268
3	1.044	17.402	69.670
4	.911	15.185	84.855
5	.693	11.549	96.404
6	.216	3.596	100.00

Tab. 7.1: PCA. Eigenvalues, total and cumulative variability for principle components.

	Component				
Dimension	1	2	3	4	5
WordCount	.593	.434	-.112	-.259	.612
Proper Nouns	-.852	.415	-.0084	-.0778	.056
Verbs	.600	.480	-.266	-.128	-.547
Adjectives	.229	-.262	.839	-.391	-.0089
Adverbs	.247	.372	.417	.787	.051
Common Nouns	.264	-.847	-.288	.224	-.0081

Tab. 7.2: PCA. Component Matrix.

the values closest to the axis are percentage of verbs and proper nouns. The difference in sign for verbs and proper nouns indicates that these variables are negatively correlated. Interestingly, according to Figure 6.27 as the proportion of proper nouns increases the proportion of verbs decreases. Also, the increase in proper nouns and decrease in verbs is significant in comparison with the variations in common nouns, adjectives and adverbs. Thus, principle component 1 reflects the relationships between proper nouns and verbs and can be used to replace this two dimensions.

The next 3 components, component 1, 2 and 3, reflect the total variability in common nouns (component 2), adjectives (component 3) and adverbs (component 4). Though, for component 4 the eigenvalue .911 from Table 7 is under 1, this value is very close 1 in comparison with component 5. Furthermore, the first four components capture 85% of all variability, which is significant and 15% more than the first three components. This suggests that component 4 is also valuable and in this case reflects the variability in adverbs.

The conclusion for this section is that using PCA we have managed to identify previously discovered in clustering relationships between proper nouns and verbs.

8. CONCLUSION

Managing and annotating personal collections of digital photographs is a difficult, tedious and boring task. However, there is a significant amount of research into this problem. This research can be divided into two main groups. The first group consists of research that uses geographical and timestamp information extracted from an image file to enable automatic management and annotation of digital photographic collections. The second group consists of research that uses unsupervised learning and computer vision techniques to create models that can attach keywords to segments of an image or to a whole image. Other research looks into ways of automatically annotating and managing digital photographs on PDAs and mobile phones.

Much effort should be dedicated to understanding user needs. The task of searching digital photo collections is overemphasised and more work is required in the direction of creating solutions useful for browsing and categorising images. Some studies into user needs suggest that users tend to search on semantics rather than image features such as colours or shapes.

This project partially relates to the problem of user needs. The aim is to understand what types of personal photographs people take and how they annotate them. An insight into the ways people annotate their personal photographs provides valuable information that can be used for constructing a meaningful and useful browsing system.

An experiment has been conducted in order to collect personal digital photographs and their annotations. There are two tools designed for the purpose of the experiment: EXIF java class library and the Collector web application. Users supplied their photographs and annotations as well as age, country of origin, gender and main interests information. 559 annotated photographs were collected from 27 different users. The 559 photographs and annotations were analysed. Also clustering using K-Means algorithm was applied to gain a further insight into the data. Three attributes the *Structure*, *Length* and *Artistic* were proposed to create new dimensions for analysis. In addition to these attributes all major

objects appeared on the photographs were recorded as well as part-of-speech information and the amount of information supplied by each annotation. Entropy measure was used to calculate the amount of information supplied by each annotation.

Based on the overall analysis of annotations and photographs the following conclusions were made. Nouns are the largest word class group and 37% of nouns are proper nouns. Adverbs appear in annotations rarely and comprise only 3% of all word classes. The large proportion of proper nouns means that the research into automatic annotation and management of digital photographic collections using geographic information is very useful. The systems described in the studies that make up this research help to discover the names of places using information such as latitude and longitude extracted from an image file and assign these names to the photographs. According to the analysis of word classes by gender, males tend to use more nouns than females and less verbs and adjectives.

The objects that appear the most frequently in the photographs consist of people, vegetation, buildings and sky. This helps to identify the types of photographs people take. At the top of the list are the photographs of people.

The main four values of the structure of the photographs are **what**, **emotion**, **who** and **location**. There are more photographs that contain **what** (inanimate objects and animate objects) than **who** (people). Moreover, 33% of annotations contain emotions of which almost half are artistic. This is an interesting property of annotations and to my knowledge it has not been explored in relation to automatic annotation. But the majority of photographs do not contain emotions. This means that the research into the automatic annotation of photographs is valuable because, for the annotations without emotions, the captions would be both satisfactory in terms of completeness and useful in terms of future referencing and browsing. Another finding is that annotations by females contain more annotations with emotions than annotations by males.

90% of annotations in the experimental dataset are of short and medium length. This means that the length of annotations is no more than 15 words. This information can be helpful in estimating the amount of storage required for annotations and the amount of words in a large collection of annotated photographs in future experiments.

Clustering provided further insight into annotations. There are 4 groups of annotations found. In the first two groups there are no annotations with emotions or artistic

annotations. In each of these groups the percentage of annotations by females and males is almost equal. The next two groups contain the majority of annotations by females. The annotations in these groups are mainly emotional and one group contains annotations that are not only emotional but also artistic.

Another interesting property observed in the discovered groups of annotation is that the percentage of proper nouns and verbs changes considerably between groups. The percentage of proper nouns decreases while the percentage of verbs increases. A large percentage of proper nouns is found in the group where annotations contain emotions and are artistic. A small percentage of proper nouns is found in the groups where annotations contain no emotions and are not artistic. Principle Component Analysis confirmed the existence of the negative correlation between proper nouns and verbs.

In the analysis of the structure of the discovered annotations, two groups of annotations contain a large percentage of **who** and **location** information. The other two groups of annotations mainly consist of annotations that describe inanimate and animate objects and not people. Interestingly, one of these groups of annotations is a group where all annotations contain emotions and are also artistic.

In the analysis of objects that relate to a particular group of annotations the following observations are made. The group with emotional and artistic annotations has the largest percentage of animals and vegetation. One of the groups that has no emotional or artistic annotations has the largest percentage of people.

It is important to mention that the dataset used for the analysis and observations produced in this report is a small dataset. The future work must include the experiment on a significantly larger scale to verify whether the conclusions and the observations still hold. The dataset should also include collecting a sufficient amount of data from individuals of different cultural backgrounds and age groups. In future work it is also important to apply other language processing techniques such as latent semantic analysis and search for attributes and algorithms useful for categorising annotations.

BIBLIOGRAPHY

- S. Arnold. Cut & paste 3-way image slideshow. URL <http://www.javascriptkit.com/script/script2/3slide.shtml>. July 2005.
- M. Balabanovic, L.L. Chu, and G.J. Wolff. Storytelling with digital photographs. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 564–571, New York, NY, USA, 2000. ACM Press. ISBN 1-58113-216-6. doi: <http://doi.acm.org/10.1145/332040.332505>.
- K. Barnard, P. Duygulu, and D.A. Forsyth. Clustering art. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages II:434–441, 2001. URL <http://kobus.ca/research/publications/CVPR-01/index.html>.
- K. Barnard, P. Duygulu, D.A. Forsyth, N. de Freitas, D.M. Blei, and M.I. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003a.
- K. Barnard, P. Duygulu, J.F.G. de Freitas, and D.A. Forsyth. Object recognition as machine translation: Exploiting image database clustering models. Unpublished manuscript, University of California at Berkeley, 2003b.
- K. Barnard, M. Johnson, and D.A. Forsyth. Word sense disambiguation with pictures. In Regina Barzilay, Ehud Reiter, and Jeffrey Mark Siskind, editors, *HLT-NAACL workshop on learning word meaning from non-linguistic data*, pages 1–5, 2003c. URL <http://kobus.ca/research/publications/LWM-03/index.html>.
- R.K. Belew. *Finding Out About*. Cambridge University Press, 2000.
- J. Berger. *Ways of Seeing*. the Penguin Group, 1972.
- E. Brill. Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging. *Comput. Linguist.*, 21(4):543–565, 1995. ISSN 0891-2017.

- D. Crouch and N. Lubben. *Visual culture and tourism*. Berg, 2003.
- D. Deriu. *Picture Essay: Souvenir Bangkok*. Berg, 2003.
- dmitri_don. Java forums - creating thumbnail from jpeg. URL <http://forum.java.sun.com/thread.jspa?threadID=223186&messageID=785701>. February 2002.
- P. Duygulu, K. Barnard, J.F.G de Freitas, and D.A Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *ECCV (4)*, pages 97–112, 2002.
- J. Edwards, R. White, and D.A. Forsyth. Words and pictures in the news. In *Proceedings of the HLT-NAACL03 Workshop on Learning Word Meaning from Non-Linguistic Data*, 2003.
- Exif. Exif.org - exif and related resources. Exif.org, 2005. URL <http://www.exif.com/>.
- A. Field. Factor analysis using spss. URL <http://www.sussex.ac.uk/Users/andyf/teaching/rm2/factor.pdf>. June 2004.
- D.A. Forsyth. Benchmarks for storage and retrieval in multimedia databases. In *Proceedings of Spie - The International Society for Optical Engineering*, volume 4676, pages 240–247. SPIE - The International Society of Optical Engineering, 2001. URL citeseer.ist.psu.edu/661295.html.
- GeoSpatial Experts. Geospatial experts link digital camera phones to gps. Gapilo Pro G3, October 2004. URL <http://www.geospatialexperts.com>.
- S. Ghadirian. Readingenglish.net - software. URL <http://www.readingenglish.net/software/>. 2004.
- D. Gussow. New lens on war. St. Petersburg Times online, May 2004. URL <http://www.sptimes.com/>.
- E. Hamilton. *JPEG File Interchange Format. Version 1.02*. C-Cube Microsystems, 1992. URL <http://www.jpeg.org/public/jfif.pdf>.
- J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2001.

- S. Harada, M. Naaman, Y.J. Song, Q Wang, and A. Paepcke. Lost in memories: interacting with photo collections on pdas. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, pages 325–333, New York, NY, USA, 2004. ACM Press. ISBN 1-58113-832-6. doi: <http://doi.acm.org/10.1145/996350.996425>.
- JavaZOOM. java upload bean. URL <http://www.javazoom.net/jzservlets/uploadbean/up1> July 2005.
- JEITA. *JEITA CP-3451. Exchangeable image file format for digital still cameras: Exif Version 2.2*. Japan Electronics and Information Technology Industries Association, April 2002. URL <http://www.exif.org/Exif2-2.PDF>. Technical Standardization Committee on AV & IT Storage Systems and Equipment.
- Kodak. Kodak professional dcs digital. Kodak, 2004. URL <http://www.kodak.com/>.
- O. Liechti and T. Ichikawa. A digital photography framework supporting social interaction and affective awareness. In *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*, pages 186–192, London, UK, 1999. Springer-Verlag. ISBN 3-540-66550-1.
- O. Liechti and T. Ichikawa. A digital photography framework enabling affective awareness in home communication. *Personal and Ubiquitous Computing*, 4(1), 2000.
- M. Lux, J. Becker, and H. Krottmaier. Caliph&Emir: Semantic Annotation and Retrieval in Personal Digital Photo Libraries. Caliph&Emir SourceForge .NET, October 2004. URL <http://caliph-emir.sourceforge.net/pdf/CaliphEmir-CAISE03.pdf>.
- P. M. Mitchell, M.A. Marcinkiewicz, and B. Santorini. Building a large annotated corpus of english: the penn treebank. *Comput. Linguist.*, 19(2):313–330, 1993. ISSN 0891-2017.
- mozart-oz.org. Penn tagset. URL <http://www.mozart-oz.org/mogul/doc/lager/brill-tagged> July 2004.
- M. Naaman, S. Harada, Q. Wang, and A. Paepcke. Adventures in space and time: Browsing personal collections of geo-referenced digital photographs. Technical Report, Stanford University, 2004a.

- M. Naaman, A. Paepcke, and H. Garcia-Molina. From where to what: Metadata sharing for digital photographs with geographic coordinates. Technical Report, Stanford University, June 2004b.
- M. Naaman, Y.J. Song, A. Paepcke, and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, pages 53–62, New York, NY, USA, 2004c. ACM Press. ISBN 1-58113-832-6. doi: <http://doi.acm.org/10.1145/996350.996366>.
- M. Naaman, Y.J. Song, A. Paepcke, and H. Garcia-Molina. Automatically generating metadata for digital photographs with geographic coordinates. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pages 244–245, New York, NY, USA, 2004d. ACM Press. ISBN 1-58113-912-8. doi: <http://doi.acm.org/10.1145/1013367.1013417>.
- Nikon. D2x. Nikon, 2004. URL <http://www.europe-nikon.com/>.
- PostgreSQL Global Development Group. Postgresql: The world’s most advanced open source database. URL <http://www.postgresql.org/>. September 2005.
- D. Sandle. Joe’s bar, douglas, isle of man: Photographic representations of holidaymakers in the 1950s. In D. Deriu, editor, *Visual culture and tourism*. Berg, 2003.
- J.C. Scherer. The photographic document: Photographs as primary data in anthropological enquiry. In E. Edwards, editor, *Anthropology and Photography 1860-1920*. Yale University Press, 1992.
- G. Seshadri. Advanced form processing using jsp. URL <http://www.javaworld.com/javaworld/jw-03-2000/jw-0331-ssj-forms.html>. March 2003.
- L.I. Smith. A tutorial on principal components analysis. URL <http://kybele.psych.cornell.edu/torial.pdf>. February 2002.
- Sun Microsystems Inc. Java technology. URL <http://java.sun.com/>. September 2005.
- T. Tachibanaya. Description of exif file format. URL <http://park2.wakwak.com/tsuruzoh/Computer/Digicams/exif-e.html>. February 2001.

-
- T. Thang. Byteconverter java class. URL <http://www.cmp.uea.ac.uk/people/researchers/a2>
March 2005.
- A. Wilhelm, Y. Takhteyev, R. Sarvas, N.V. House, and M. Davis. Photo annotation on a camera phone. In *CHI '04 extended abstracts on Human factors in computing systems*, pages 1403–1406, New York, NY, USA, 2004. ACM Press. ISBN 1-58113-703-6. doi: <http://doi.acm.org/10.1145/985921.986075>.
- D. Zeitlyn. Visual anthropology at kent. URL <http://lucy.kent.ac.uk/VA/>. September 2003.

APPENDIX

No	Name	Description	Min	Max	Mean	Std Dev	Unique	Miss.	Type
1	AnnotationId	Annotation Identifier	1	570	285.945	164.294	559	0	NUM.DISC
2	age-group	Age group	-	-	-	-	5	0	CATEGORICAL
3	gender*	Gender (f/m)	-	-	-	-	2	0	CATEGORICAL
4	country_of_origin	Country of Origin	-	-	-	-	12	0	CATEGORICAL
5	Emotion*	Annotation contains emotions (Y/N)	0	1	0.332737	0.471193	2	0	NUM.DISC
6	What*	Annotation contains a reference to animate or inanimate objects (Y/N)	0	1	0.549195	0.497574	2	0	NUM.DISC
7	Who*	Annotation contains a reference to (Y/N)	0	1	0.31127	0.463013	2	0	NUM.DISC
8	Location*	Annotation contains a reference to locations (Y/N)	0	1	0.221825	0.415474	2	0	NUM.DISC
9	Event*	Annotation contains a reference to events	0	1	0.110912	0.314024	2	0	NUM.DISC
10	Timeline*	Annotation contains date or times	0	1	0.0572451	0.23231	2	0	NUM.DISC
11	Action*	Annotation contain reference to actions	0	1	0.0143113	0.118771	2	0	NUM.DISC
12	Length*	Length of annotation	-	-	-	-	3	0	CATEGORICAL
13	ProperNouns*	Count of proper nouns	0	7	1.12165	1.26726	8	0	NUM.DISC
14	Verbs*	Count of verbs	0	15	0.855098	1.71245	13	0	NUM.DISC
15	Adjectives*	Count of adjectives	0	11	0.504472	0.934088	8	0	NUM.DISC
16	Adverbs*	Count of adverbs	0	4	0.227191	0.627553	5	0	NUM.DISC
17	Other*	Count of word classes that are not nouns, adverbs, adjectives or verbs	0	23	2.7263	3.5531	21	0	NUM.DISC
18	CommonNouns*	Count of common nouns	0	12	1.85331	2.03369	13	0	NUM.DISC
19	Nouns*	Count of nouns	0	15	2.97496	2.5185	16	0	NUM.DISC
20	Artistic*	Annotation is artistic (Y/N)	0	1	0.159213	0.365874	2	0	NUM.DISC
21	NormalisedInfo*	The amount of information in annotation (measured in bits)	0	0.477133	0.0618273	0.0618803	349	0	NUM.CONT.
22	WordCount*	Count of words in annotation	1	53	7.28801	7.97342	40	0	NUM.DISC

Tab. .1: Data Dictionary.

Page	Java Bean	Scope	Alias
members/index.jsp	pfa.Manager	session	manager
	pfa.Author	session	author
members/slideshow.jsp	pfa.Manager	session	manager
	pfa.Collection	page	col
	pfa.Photo	page	
members/show_all_collections.jsp	pfa.Manager	session	manager
members/file_info.jsp	pfa.ExifInfo	page	exif
	pfa.Photo	session	
members/edit_photograph.jsp	pfa.Manager	session	manager
members/details.jsp	pfa.Manager	session	manager
	pfa.Author	page	author
members/add_photograph.jsp	pfa.Manager	session	manager
	pfa.PhotoUpload	request	photoUpload
	javazoom.upload.UploadBean	page	upBean
	analyser.Exif	page	exifRW
members/collection_photos.jsp	pfa.Manager	session	manager
	pfa.Collection	session	userCollection
members/add_collection.jsp	pfa.Manager	session	manager
	pfa.Collection	page	collection
admin/report.jsp	pfa.Report	session	report
	pfa.Collection	page	col
	pfa.Photo	page	photo
registration/process_form.jsp	pfa.Registration	request	formHandler

Tab. .2: Java Pages that use Java Beans.

Country Name
Afghanistan
Albania
Algeria
American Samoa
Andorra
Angola
Anguilla
Antarctica
Antigua and Barbuda
Argentina
Armenia
Aruba
Ascension Island
Australia
Austria
Azerbaijan
Bahamas
Bahrain
Bangladesh
Barbados
Belarus
Belgium
Belize
Benin
Bermuda
Bhutan
Bolivia
Bosnia and Herzegovina
Botswana
Bouvet Island
Brazil
British Indian Ocean Territory
Brunei Darussalam
Bulgaria
Burkina Faso
Burundi
Cambodia
Cameroon
Canada
Cape Verde
Cayman Islands
Central African Republic
Chad
Chile

Country Name
China
Christmas Island
Cocos (Keeling) Islands
Colombia
Comoros
Democratic Republic of the Congo (Kinshasa)
Congo, Republic of (Brazzaville)
Cook Islands
Costa Rica
Ivory Coast
Croatia
Cuba
Cyprus
Czech Republic
Denmark
Djibouti
Dominica
Dominican Republic
East Timor Timor-Leste
Ecuador
Egypt
El Salvador
Equatorial Guinea
Eritrea
Estonia
Ethiopia
Falkland Islands
Faroe Islands
Fiji
Finland
France
French Guiana
French Metropolitan
French Polynesia
French Southern Territories
Gabon
Gambia
Georgia
Germany
Ghana
Gibraltar
Great Britain
Greece
Greenland
Grenada
Guadeloupe

Country Name
Guam
Guatemala
Guernsey
Guinea
Guinea-Bissau
Guyana
Haiti
Heard and Mc Donald Islands
Holy See
Honduras
Hong Kong
Hungary
Iceland
India
Indonesia
Iran (Islamic Republic of)
Iraq
Ireland
Isle of Man
Israel
Italy
Jamaica
Japan
Jersey
Jordan
Kazakhstan
Kenya
Kiribati
Korea, Democratic People's Rep. (North Korea)
Korea, Republic of (South Korea)
Kuwait
Kyrgyzstan
Lao, People's Democratic Republic
Latvia
Lebanon
Lesotho
Liberia
Libya
Liechtenstein
Lithuania
Luxembourg
Macao
Macedonia
Madagascar
Malawi
Malaysia

Country Name
Maldives
Mali
Malta
Marshall Islands
Martinique
Mauritania
Mauritius
Mayotte
Mexico
Micronesia, Federal States of
Moldova, Republic of
Monaco
Mongolia
Montserrat
Morocco
Mozambique
Myanmar, Burma
Namibia
Nauru
Nepal
Netherlands
Netherlands Antilles
New Caledonia
New Zealand
Nicaragua
Niger
Nigeria
Niue
Norfolk Island
Northern Mariana Islands
Norway
Oman
Pakistan
Palau
Palestinian National Authority
Panama
Papua New Guinea
Paraguay
Peru
Philippines
Pitcairn Island
Poland
Portugal
Puerto Rico
Qatar
Reunion Island

Country Name
Romania
Russian Federation
Rwanda
Saint Kitts and Nevis
Saint Lucia
Saint Vincent and the Grenadines
Samoa
San Marino
Sao Tome and Principe
Saudi Arabia
Senegal
Serbia and Montenegro
Seychelles
Sierra Leone
Singapore
Slovakia (Slovak Republic)
Slovenia
Solomon Islands
Somalia
South Africa
South Georgia and South Sandwich Islands
Spain
Sri Lanka
Saint Helena
St. Pierre and Miquelon
Sudan
Suriname
Svalbard and Jan Mayen Islands
Swaziland
Sweden
Switzerland
Syria, Syrian Arab Republic
Taiwan, Republic of China
Tajikistan
Tanzania
Thailand
Tibet
Timor-Leste (East Timor)
Togo
Tokelau
Tonga
Trinidad and Tobago
Tunisia
Turkey
Turkmenistan
Turks and Caicos Islands

Country Name
Tuvalu
Uganda
Ukraine
United Arab Emirates
United Kingdom
United States
U.S. Minor Outlying Islands
Uruguay
Uzbekistan
Vanuatu
Vatican City State (Holy See)
Venezuela
Vietnam
Virgin Islands (British)
Virgin Islands (U.S.)
Wallis and Futuna Islands
Western Sahara
Yemen
Zaire
Zambia
Zimbabwe

Tab. .3: Country of Origin.

Age Group
under 12
12 - 18
19 - 25
26 - 35
36 - 45
46 - 55
55 - 65
65 - 75
over 75

Tab. .4: Age groups.

Structure	Count	Percentage
what	173	31.28%
emotion,what	60	10.85%
who	55	9.95%
emotion	48	8.68%
location,who	24	4.34%
emotion,who	23	4.16%
emotion,location,who	19	3.44%
location,what	19	3.44%
emotion,location,what	12	2.17%
event	11	1.99%
location	10	1.81%
event,who	9	1.63%
event,location,who	8	1.45%
event,timeline	8	1.45%
what,who	7	1.27%
event,location	6	1.08%
timeline,what	5	0.90%
action,location,who	4	0.72%
emotion,event,who	4	0.72%
event,what	4	0.72%
location,what,who	4	0.72%
emotion,timeline,what	3	0.54%
emotion,what,who	3	0.54%
event,location,what,who	3	0.54%
location,timeline,what	3	0.54%
emotion,event,what	2	0.36%
emotion,event	2	0.36%
emotion,location	2	0.36%
emotion,location,timeline	2	0.36%
emotion,timeline,who	2	0.36%
timeline,what,who	2	0.36%
emotion,location,what,who	1	0.18%
action	1	0.18%
action,emotion,who	1	0.18%
action,what,who	1	0.18%
action,who	1	0.18%

Structure	Count	Percentage
emotion,event,location	1	0.18%
emotion,location,who,timeline	1	0.18%
event,location,timeline	1	0.18%
event,location,timeline,who	1	0.18%
event,location,what	1	0.18%
event,timeline,what	1	0.18%
event,timeline,what,who	1	0.18%
event,timeline,who	1	0.18%
location,timeline	1	0.18%
location,timeline,who	1	0.18%
timeline	1	0.18%

Tab. .5: Distinct values of structure attribute with count and percentage.

Object Category	Count	Percentage
people	331	24.78%
vegetation	218	16.32%
buildings	129	9.66%
sky	118	8.83%
constructed indoor objects	78	5.84%
animals	64	4.79%
household objects	64	4.79%
landscape	62	4.64%
constructed outdoor objects	56	4.19%
water	54	4.04%
transport objects	52	3.89%
works of art	37	2.77%
man made objects	25	1.87%
food	25	1.87%
other	19	1.42%
clothes	4	0.30%

Tab. .6: Distinct values of categories of objects with count.

distance to centroid of cluster	3 clusters	4 clusters	5 clusters
1	1.289456	1.589686	1.288661
2	1.44566	1.123262	1.657985
3	0	1.29332	1.71659
4		0	1.890354
5			0

Tab. .7: Distance between cluster centroids.

Categorised Objects	Cluster1	%age	Cluster2	%age	Cluster3	%age	Cluster4	%age
people	54	19%	47	26%	108	19%	121	39%
vegetation	77	27%	32	18%	75	14%	34	11%
buildings	21	7%	20	11%	62	11%	26	8%
sky	21	7%	23	13%	50	9%	24	8%
constructed indoor objects	15	5%	5	3%	33	6%	25	8%
household objects	15	5%	4	2%	24	4%	21	7%
animals	28	10%	3	2%	31	6%	2	1%
landscape	8	3%	15	8%	31	6%	8	3%
constructed outdoor objects	15	5%	12	7%	22	4%	7	2%
water	8	3%	9	5%	24	4%	13	4%
transport objects	12	4%	6	3%	26	5%	8	3%
works of art	4	1%	0	0%	30	5%	3	1%
food	2	1%	1	1%	18	3%	4	1%
man made objects	6	2%	4	2%	8	1%	7	2%
other	2	1%	1	1%	9	2%	7	2%
clothes	0	0%	0	0%	4	1%	0	0%

Tab. .8: Distinct values of objects with count and percentage per cluster.