Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002: 53-60

Preliminary Results from the FILTER Image Categorisation and Description Exercise

Jill Evans
Institute for Learning and Research Technology, University of Bristol
jill.evans@bristol.ac.uk

Paul Shabajee
Graduate School of Education and Institute for Learning
and Research Technology, University of Bristol
paul.shabajee@bristol.ac.uk

Abstract

Although there are now vast numbers of digital images available via the Web, it is still the case that not enough is known or understood about how humans perceive and recognise image content, and use human language terms as a basis for retrieving and selecting images. There is an increasing belief that the difficulties of image management and description should be led and defined by the needs of users and by their information seeking behaviours. The Focusing Images for Learning and Teaching – an Enriched Resource (FILTER) project is investigating, through an online image description exercise, the ways that users describe different types of images. Through analysis of the exercise results, FILTER hopes to obtain an understanding of the ways in which people describe images and the factors that influence their approaches to image description and thus appropriate metadata. Preliminary analysis of data indicates that there is little consensus on the use of terms for image description or on categorisation of images into 'types'.

1. Introduction: Describing and Retrieving Images

As all forms of communication are increasingly transferred via the grammar of the visual, humans are becoming more sophisticated in their ability to recognise and interpret visual meaning and are using visual information to enhance social, cultural or learning activities [1]. The huge global financial investment in the digitisation of analogue images means that there are now immense numbers of diverse image collections available on the Web, and appetite for consumption of visual information continues to grow. However, it is still the case that not enough is known or understood about how humans perceive and recognise image content, and use human language terms as a basis for retrieving and selecting images in vast, complex, heterogeneous online environments [2]. There is a gap in understanding of how humans verbalise visual perceptions and beliefs within paradigms typically associated with text. Consequently, we cannot be sure that the image resources we are creating within the constraints of orthodox description models are actually being found, accessed and used by our target audiences.

The creators of digital image collections, in most cases, intend their collections to be used as widely and as effectively as possible. Therefore, adding as comprehensive and rich a layer of metadata as possible is essential. Metadata is particularly important for images, as without accurate description there cannot be accurate retrieval [3]. An enormous problem, however, is that images, by their very nature – that they are pictorial representations rather than verbal – are difficult to classify in orthodox textual terms. O'Connor [4] notes that images are complex '… by being literally and figuratively unquotable, everlastingly slipping through in the instance of being identified, seized for scrutiny'. Additionally, '… there is no saying just how many words or just which words are required to describe any individual pictures' [5]. Typically, therefore, visual material requires description of a greater depth than textual resources in order to convey an understanding of its complexity and the implicit multiple layers of meaning and content [6].

Interpreting the meaning or essence of an image and translating that into a surrogate form for retrieval and management is a complicated and diffi-

cult process. Description of images is made complex because of the impossibility of agreement on the meaning, or on the most critical aspects represented. Meaning will be constructed differently, and various elements will have greater relevance to different users depending on, for example: the intended use for the image, the user's area of study or research, the user's educational level of ability, his/her level of cultural, social and historical awareness, and so on. Any or all of these factors will provoke a different user reaction to the image and influence perception of what is depicted and what is expressed; clearly, the vocabularies employed by this wide range of users will all be quite different and will be impossible to anticipate by the providers of the digital image collection.

For example, the following image is a histological stain of tissue from a human gland.

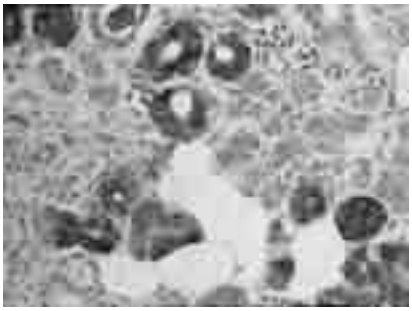


Image 1. (Copyright the Bristol Biomedical Image Archive)

The intended audience is doctors, nurses, medical students, and so on. However, the image could equally be of use to an art student looking for inspiration for a textile design. Evidently, the subject content of an image can hold a variety of meanings depending on the primary need of the user but it would be impractical and almost certainly impossible for the digital image collection provider to attempt to describe and classify all these possible contexts and uses.

As Conniss, Ashford and Graham [2], have observed the depth and manner in which an image is catalogued will have consequences for its ability to be retrieved by a potential user. They go on to note that, although comprehensive, in-depth cataloguing can provide multiple and varied entry points for diverse audiences, this level of description is in practice a very time-intensive and ultimately costly process. If images are to be added to a collection regularly, the issue of sustaining such a cataloguing model needs to be addressed.

A book or journal has a generally predictable structured format. Subject indexing of textual materials is usually aided by the availability of several sources of information: preface, title, table of contents, and so on, from which to determine the primary content, aims, scope and purpose. It is normally possible to locate appropriate subject headings to describe the item from within a controlled vocabulary or thesaurus. Cataloguers of visual material typically have no such standard, recognised sources at their disposal, and images can take many varied forms. Howard Besser [7], on the topic of image description writes:

(of a book) '... Authors and publishers go to great lengths to tell us what this purpose is, citing it in the preface and introduction, on dust covers, and often on the book's cover. Images do not do this. To paraphrase one prominent author speaking of museum objects, unlike a book, an image makes no attempt to tell us what it is about. Even though the person who captured an image or created an object may have had a specific purpose in mind, the image or object is left to stand on its own and is often used for purposes not anticipated by the original creator or capturer'.

There will usually be little textual information accompanying the image to explain the social, political, historical, religious or cultural context in which it is embedded, or for what purpose the image was created. The image cataloguer may, therefore, have to invest considerable time in research in order to answer these questions before attempting image description.

Most images comprise two aspects of meaning: what they are 'of', that is, what they depict, such as an identifiable person, object or place; what they are 'about', an underlying theme, emotion, message or other abstract concept that is extracted by interpretation. These 'ofness' and 'aboutness' aspects are based on the model developed by Panofsky [8] who described levels of meaning as pre-iconographic (of) and iconographic (about). Krause [9] has used the terms 'hard indexing' and 'soft indexing'. For example, the German Expressionist artist Kathe Kollwitz's charcoal drawing overtly depicts an impoverished mother and her children, but is covertly about the plight of the working classes during the years of the German depression.

Image 2. (Copyright Brooklyn College History Dept.)

Far from being a straightforward depiction, this image was intended as a stringent criticism of government policy and beliefs by an artist violently opposed to them. However, without an understanding of the historical and social factors that influenced the creation of this image, the meaning cannot fully be understood. Cataloguing images for this type of conceptual or abstract information need is extremely complex as possibly no two indexers will reach consensus on the subjective qualities or meaning of an image. As Bradfield [10] notes:

'The problem with the retrieval of visual material is that it evokes concepts related to the reality which it represents. Such concepts are not easily expressed in words but are the 'sought' features of that visual image. Equally, the reality is not always readily expressible in words'.

It can be seen that determining the focus of an image for indexing poses a huge challenge, both from a depiction and expression perspective. A user reacts to an image on many levels in constructing a feeling for its meaning and connotations. Even images that are abstract and elusive in content are capable of evoking feelings and attempts to communicate those feelings through words. Ornager [11], based on research conducted with image users, suggests that image indexing should indicate what the image depicts (its 'ofness'), what it expresses ('aboutness'), and, additionally, the contexts in which it can be used.

It is generally accepted that the aim of image indexing is not only to provide access based on descriptions of attributes, but also to provide access to useful and logical groupings of images [12]. However, images, unlike verbal representations, share no common hierarchy or taxonomy of definitions and relationships. O'Connor [5] gives as an example a series of images of an elephant, a sheep and a horse, all of which could more broadly be described as animals. A user looking for an image of a horse might also be interested in seeing images of other animals, or of animals that graze, or of mammals, or of four-legged mammals. How can images be indexed and represented to users so as to make these complex but potentially useful relationships more visible and accessible? How can indexers predict what kind of relationships will be valuable to diverse communities of users? [13].

2. The FILTER Project

There is an increasing belief that the difficulties of image management and description should be led and defined by the needs of users and by their information seeking behaviours [6], [14], [11], [5]. The Focusing Images for Learning and Teaching – an Enriched Resource (FILTER) project (http://www.filter.ac.uk/) is investigating, through an online image description exercise (http://www.filter.ac.uk/exercise/), the ways that users describe different types of images. About 40 copyright-free images of varying original types (e.g. map, etching, drawing, chart, painting, and so on) and subject content were placed online; individuals from all aspects of tertiary education (but not restricted to these) were invited to participate in the exercise by, firstly, describing the subject content of each image in a series of unlimited keywords, and secondly, describing the type of image (the original type rather than the digital type, as all could legitimately be described as a 'photograph'). Through analysis of the exercise results, FILTER hopes to obtain an understanding of the ways in which people describe images and the factors that influence their approaches to image description. For example: are there particular 'types' of images (e.g. line drawings, graphs, maps) that are easier to describe – and where more consensus is reached? How do users react to images that are more abstract or ambivalent in content compared to images where the content is clear – are more words used to describe ambiguous content? When text is included in an image, does this influence choice of keywords? Is there a difference in the way users from different subject areas approach image description?

FILTER is working with academic image users to develop an exemplar database of image-based learning and teaching materials that demonstrate effective use of images across subject areas and in a range of pedagogical contexts. FILTER has recognised that there are complex issues involved in making these materials and the images embedded within them available in a heterogeneous environment [15]. Both resources and images need to be described and represented in such a way as to encourage users from a specific subject area to look beyond that to examples of image use in other disciplines, which might be relevant. In order to achieve this cross-searching and potential transference of knowledge and expertise, we first need to achieve an understanding of how people perceive and describe images.

3. The Image Exercise

It was essential that the images included should be very diverse in 'type', subject content and style of content representation (i.e. degree of clarity/ambiguity). Also important was that the range of images should be typical of those used in different pedagogical contexts. Images were randomly assigned a number from 1-41. On accessing the Web page, participants were presented with a random image accompanied by a questionnaire. Once the questionnaire had been completed and submitted for that image, participants were offered the choice of proceeding to the next image in the sequence. Participants were not permitted to opt out of 'difficult' images or choose which images to describe.

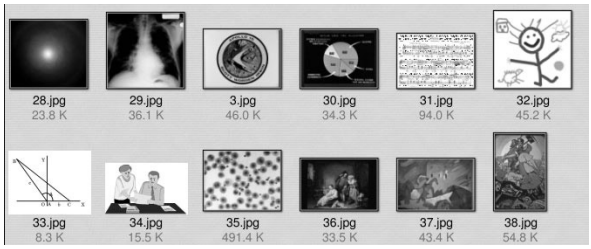


Image 3. Screenshot of a selection of images used in the exercise (Image copyright NASA, NOAA, FWS, Brooklyn College History Dept., AICT, Bristol BioMed, FILTER)

3.1 The Questionnaire

In addition to adding their descriptions, participants were required to add their area of study, teaching or research and their status (e.g. higher education lecturer, librarian, further education student). The questionnaire was by default anonymous but participants were given the option of adding their contact details should they wish to be involved in future FILTER research (UK Data Protection regulations were a consideration here). From this data we can identify 251 individuals (Table 1). Participants study, work or teach within a diverse range of subject areas, for example: Education, Music, Art and Design, Environmental Science, Medicine, History, Language and Literature, IT, Biology, Psychology, Engineering, Librarianship, Archaeology, Law, Business, Management, and so on.

3.2 The Sample

The sample for the exercise was self-selecting but particular groups of potential image users were targeted for publicity. Information about the exercise was sent to a variety of UK and international mailing lists in the education, library, imaging and educational technology fields. The exercise was also published on relevant Web sites, in newsletters and promoted by word of mouth.

Table 1. Status of participants and numbers participating

Status of participant	Number participating
Higher Education lecturer	62
Further Education lecturer	7
Higher Education student	16
Further Education student	5
Researcher	43
Librarian (HE & FE)	68
Educational technologist	18
School teacher	3
Administrative staff	8
Technician	3
Digitisation staff	4
Other categories	3
Status not given	11

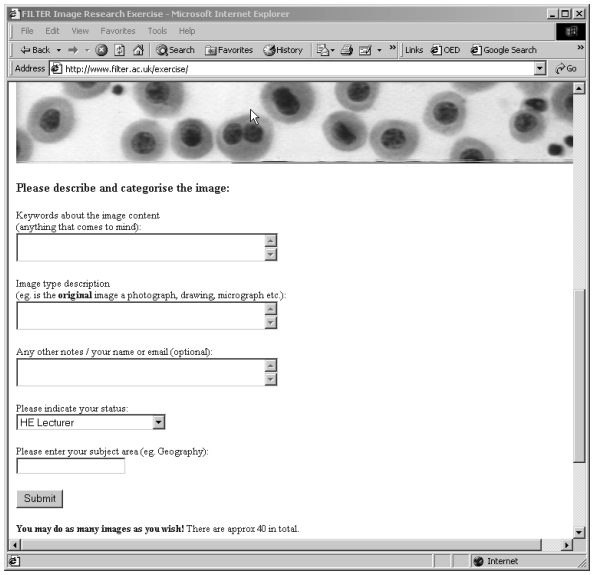


Image 4. Screenshot of questionnaire

4. Preliminary Results

The aim of the survey is to gain an initial insight into the ways a variety of image users perceive, describe and categorise images. The survey is not intended to be a definitive study in the area, but rather to highlight issues for further, more rigorous, investigation and research. The self-selecting nature of the sample and the self-reporting format of the Web questionnaire do raise questions of validity and unreliability: it is impossible to ascertain to what degree respondents are representative of the community of image users, or whether a full range of expertise in image use – from novice to expert – is present. However, as a basic snapshot of the current level of expertise in the field of visual information description and classification across multiple disciplines, we believe the survey has great potential relevance. The substantial proportion of respondents who provided their contact details indicates that preliminary findings can be followed up via interview or other in-depth questioning.

4.1 Analysis

An initial analysis of the data has been conducted by taking the submissions from the Web form, entering the data into a relational database and querying the database directly or producing output for more detailed analysis in the statistical analysis package SPSS and qualitative analysis software ATLAS-ti

where appropriate. At the time of this preliminary analysis there had been 1150 responses.

The 'Other' category contains a range of roles which, in cases where respondents provided additional details, we have been able to further categorise (see Table 1).

Figure 2 shows that the number of responses for each image was fairly evenly distributed as expected due to the random presentation of the images to the participants.

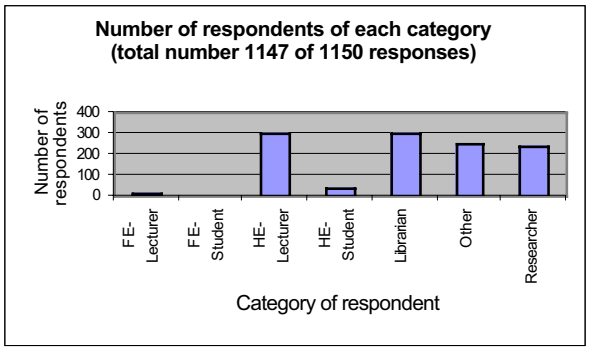


Figure 1. Number of respondents of each category

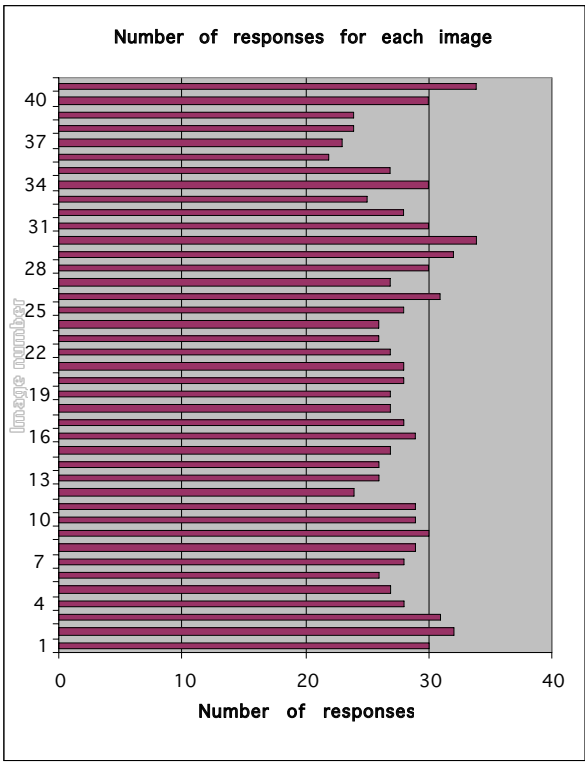
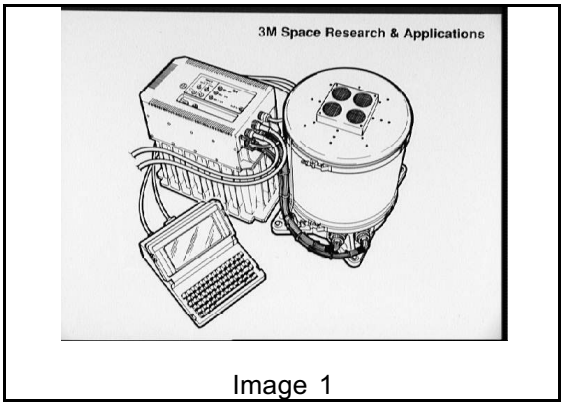


Figure 2. Number of responses for each image

Table 2. Terms used to describe the type of image for Image 1 (Image copyright NASA)



Term or Phrase	Number of times used
drawing	16
line drawing	2
technical drawing	1
artwork	1
technical sketch of overall arrangement of kit	1
sketch	1
pen-and-ink drawing on paper	1
line illustration for technical manual	1
engineering drawing	1
drawing probable computer generated	1
computer-generated image	1
computer generated image	1
two tone drawing	1

4.2 Categorisation of Image Types

A total of 391 terms and phrases were used to categorise the types of image in the exercise. This number includes terms and phrases with obvious spelling errors and hyphenated alternatives.

Table 3 shows the terms and phrases used more than once to describe the type of image and the number of times it was used across the whole collection. This illustrates a number of interesting points: 1) it seems that synonyms occur frequently, e.g. photo/photograph and computer image/computer-generated image/digital image 2) qualifiers are used extensively, e.g. colour, black and white, computer generated, annotated 3) misspellings occur, e.g. 'satelite'. This is more common than shown here as many spelling mistakes occurred only once 4) some contextual terms are used, e.g. scientific, and 5) participants use terms such as 'don't know' and 'probable' as indicators that they are unsure how to categorise the image. We plan to use this data in combination with earlier work to develop a set of vocabularies and relationships that model how participants of this and other studies have described image types.

Table 3. Terms and phrases used more than once to describe the type of image and the number of times it was used (for the whole set of images)

Term or Phrase	N	Term or Phrase	N
photograph	23	annotated map	2
drawing	21	colour drawing	2
photo	17	computer generated drawing	2
painting	10	computer generated image	2
colour photograph	8	computer image	2
print	7	don't know	2
diagram	6	engraving	2
black and white	5	figure	2
chart	5	logo	2
micrograph	5	map drawing	2
sketch	5	microscope image	2
black and white photograph	4	microscopic image	2
computer-generated image	4	photograph (colour)	2
graph	4	photograph of graphic	2
illustration	4	photographs	2
line drawing	4	picture	2
artwork	3	printed map	2
b&w photograph	3	printout	2
cartoon	3	satelite image	2
computer generated	3	scan	2
lithograph	3	scanned image	2
map	3	scientific graph/ diagram	2
microscope slide	3	technical drawing	2
photo of chart	3	trace	2
aerial or satellite photograph	2	transparency	2
aerial photograph	2		

For each image a clear pattern emerged, as illustrated in Table 2 (for the case of image 1). A small number (generally 1-4) of terms or phrases were used by a large proportion of participants, and a larger number of terms were used only once, i.e. by only one participant in each case. This implies that for each image there is a small set of 'image types' that are most commonly used.

Figure 3 shows the number of distinct terms or phrases used to describe each image type in the collection. Initial analysis indicates that there are underlying patterns in this data. For example, images 6, 11, 18 and 26 all have 'number of types of image' of 5 or less and in each case the most popular term to describe that image type is 'photograph', whilst this was not the case for any of those images with more than 15 terms or phrases.

A more detailed analysis will, we hope, provide insight into exactly what characteristics of an image participants have focused on in order to arrive at a categorisation type.

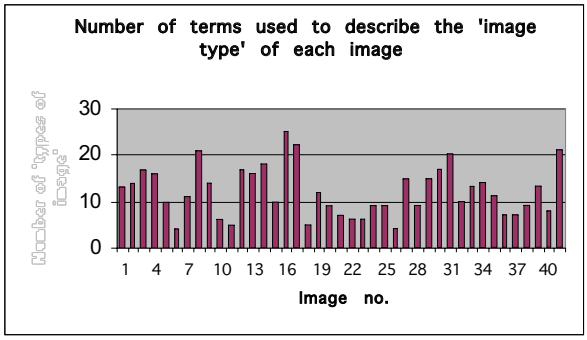


Figure 3. Number of terms used to describe the image type of each image

4.3 Describing the Images

The description of images is more complex to analyse. Initial analysis described here uses only individual words although participants frequently used prose to describe the images. Thus the meanings that can be derived from that prose are lost in this analysis. Figure 4 shows the number of terms used per image. For all images it appears that the distribution of the number of terms is very flat. Figure 5 shows the number of times a word was used to describe image number 1.

For image 1 there are 130 terms in total used by 30 participants; the most commonly used is 'computer', used 18 times, followed by 'equipment', used 11 times. There are 54 terms that have been used only once. This pattern is similar to that for the image type data above but with many more terms and a longer tail. It is more problematic to analyse these with respect to categorising the words and thus identifying common themes. This more detailed analysis will be conducted over the next months, however, we describe our initial analysis and findings below.

We are currently analysing the descriptive terms used for each image in detail. Table 2 shows the words used to describe image 1. Figure 6 shows patterns of the co-occurrence of words in descriptions for image 1. The terms joined by lines were terms that co-occurred more than once in the descriptions and the thick lines indicate those that co-occurred more than three times. Clearly, for image 1 the dominant concepts are: computer, equipment, control/controlled, space and research.

Preliminary findings indicates that, on further analysis, it should be possible to categorise the words and phrases in terms of the nature of the description, for instance: shape of content, colour, object, person, historical/temporal. We should also be able to judge the extent to which the inclusion of text in an image influences amount and choice of words. Such analysis will, we believe, provide a means of categorising the characteristics in the images that participants have used to describe the images and of gaining insight into the elements or facets of the images that they use to make their choices of terms.

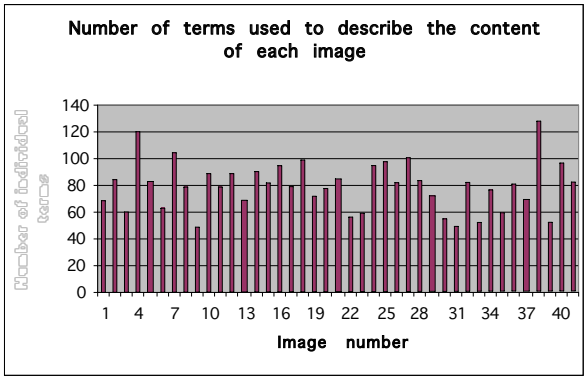


Figure 4. Number of terms used to describe the content of each image

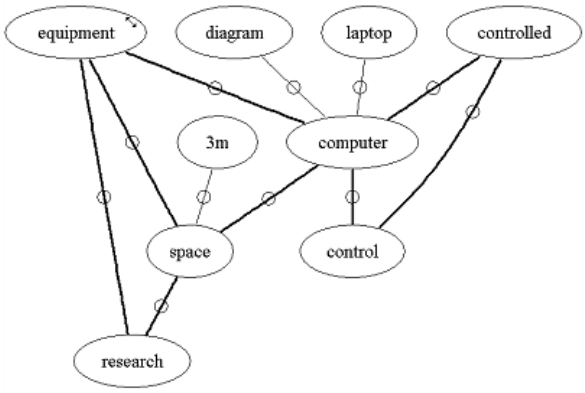


Figure 6. Graph of co-occurrences for image 1 produced with Graphvis software [16]

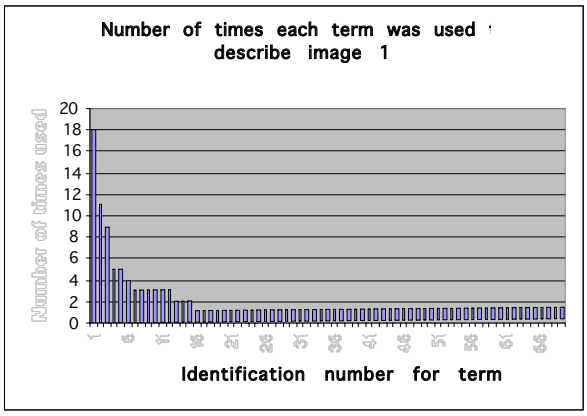


Figure 5. Number of times each term was used to describe image 1

5. Discussion and Moving Forward

The findings, even for this early analysis, seem to point to the advantages of, and need for, the use of controlled vocabularies and thesauri to overcome common spelling variations and errors. Current image metadata standards do not seem to provide vocabularies for representing the full range of 'image type' descriptors used by the participants. Of course, it is not clear that such a categorisation is necessary, or would be beneficial. However, if a digital image collection contains tens of thousands of images of maps, and a user is specifically seeking paintings of maps, then certainly some form of metadata standards would be required to enable such a search to be carried out efficiently.

On the basis of the data presented in this paper, one possible structure for an image type categorisation system would be to define core types (e.g. photograph, painting, map), with an optional set of qualifiers for providing additional meaning based on visual aspects of the image. These might be, for example, colour/black & white, mode of generation/technique/process, angle/view, scale, spatial/temporal. Clearly there is some potential here for the development of a refined set of qualifiers for the DCMI Type Vocabulary [17].

The Getty Art and Architecture Thesaurus (http://www.getty.edu/research/tools/vocabulary/aat/) is a highly authoritative vocabulary in the broad arts fields; FILTER suggests that it would be useful to develop a broad categorisation of image types that could be used across a diverse range of subject areas and that is specifically focused on the needs of the tertiary education community. The results of the exercise clearly demonstrate that there is a lack of consensus on the process of recognising and categorising images by type. It is possible that this lack of agreement acts as a barrier in the successful retrieval and use of images and that this issue could be addressed by the establishment of a common vocabulary.

There is clearly very large diversity in the terms used to describe the content of the images in this sample. While dominant terms exist for each image, they represent only a small percentage of the terms used and are not used by all participants. While thesauri can clearly facilitate the mapping of terms to any descriptive metadata, it is not clear from our current analysis that such a mapping would meet the needs of the participants. Once performed, these analyses should help to shed light on this issue.

Other analyses that we are conducting are based on the identification of patterns in the data related to, for instance, the role or "subject area" of the participant and their classification of both image type and content. For example, we might hypothesise that participants from different subjects will describe the content in different ways. It is hoped that this work will enable FILTER and other projects develop effective means of helping those developing online image collections for educational purposes to provide their users with tools to retrieve images.

References

[1] Kress, G. and van Leeuwen, T., *Reading images: the grammar of visual design*. Routledge. London, 1996.

[2] Conniss, L.R., Ashford, A.J. and Graham, M.E., *Information seeking behaviour in image retrieval: VISOR I final report*. (Library and Information Commission Research Report: 95), Newcastle: Institute for Image Data Research, 2000.

[3] Rorvig, M.E., Turner, C.H. and Moncada, J., "The NASA image collection visual thesaurus", *Journal of the American Society for Information Science*, 50 (9), 1999, 794-798.

[4] O' Connor, B.C., "Access to moving image documents: background concepts and proposals for surrogates for films", *Journal of Documentation*, 41 (4), 1985, 209-220.

[5] O' Connor, B.C., "Pictures, aboutness, and user-generated descriptors", *SIGVIS Newsletter* [online], 1 (2). 1996, Available from: http://www.asis.org/SIG/SIGVIS/b_oconnor.html [Accessed 28 June 2002].

[6] Turner, J.M., "A typology for visual collection", *Bulletin of the American Society for Information Science* [online], 25 (6). 1999. Available from: http://www.asis.org/Bulletin/Aug-99/turner.html [Accessed 28 June 2002].

[7] Besser, H., Visual access to visual images: the UC Berkeley Image Database Project. *Library Trends*, 38 (4), Spring, 1990, 787-798.

[8] Panofsky, E., "*Meaning in the visual arts: papers in and on art history*", Garden City, N.Y.: Doubleday Anchor, 1955.

[9] Krause, M. G., "Intellectual problems of indexing picture collections", *Audiovisual Librarian*, 14 (2), 1988, 73-81.

[10] Bradfield, V., "Slides and their users: thoughts following a survey of some slide collections in Britain", *Art Libraries Journal*, Autumn, 1977, 4-21.

[11] Ornager, S., "Image retrieval: theoretical analysis and empirical user studies on accessing information in images", *Proceedings of the 60th ASIS Annual Meeting*, 34, 1997, 202-211.

[12] Layne, S. S., Some issues in the indexing of images. *Journal of the American Society for Information Science*, 45 (8), 1994, 583-588.

[13] Shabajee, P., "Primary multimedia objects and educational metadata: a fundamental dilemma for developers of multimedia archives", *D-Lib Magazine* [online], 8 (6). 2002, Available from: http://www.dlib.org/dlib/june02/shabajee/06shabajee.html, [Accessed 28 June 2002].

[14] Jorgensen, C., "Attributes of images in describing tasks", *Information Processing & Management*, 34 (2/3), 1997, 161-174.

[15] Evans, J., Conole, G. and Youngs, K., "FILTER: Focusing Images for Learning and Teaching - an Enriched Resource". In: Kennedy, G. et al., eds., *Meeting at the crossroads: proceedings of the 18th annual conference of the Australian Society for Computers in Learning and Tertiary Education (ASCILITE), Melbourne, 9-12 December 2001*. Melbourne: The University of Melbourne, 2001, 189-196.

[16] AT&T, *Graphviz – open source graph drawing software* [online]. 2002, Available from: http://www.research.att.com/sw/tools/graphviz/ [Accessed 28 June 2002].

[17] DCMI. "*DCMI Type Vocabulary*". Available from: http://dublincore.org/documents/2000/07/11/dcmi-type-vocabulary/ [Accessed 28 June 2002].