Conference paper

Clough, P., Joho, H. and Sanderson, M. (2005) *Automatically organising images using concept hierarchies.* In: Multimedia Information Retrieval Workshop 2005 in conjunction with the 28th annual ACM SIGIR conference on Information Retrieval, August 19 2005, Salvador, Brazil.

Automatically Organising Images using Concept Hierarchies

Paul Clough
University of Sheffield
Western Bank
Sheffield, UK
+44 114 222 2664
p.d.clough@sheffield.ac.uk

Hideo Joho
University of Sheffield
Western Bank
Sheffield, UK
+44 114 222 2664
h.joho@sheffield.ac.uk

Mark Sanderson
University of Sheffield
Western Bank
Sheffield, UK
+44 114 222 2648
m.sanderson@sheffield.ac.uk

ABSTRACT

In this paper we discuss the use of concept hierarchies, an approach to automatically organize a set of documents based upon a set of concepts derived from the documents themselves for image retrieval. Co-occurrence between terms associated with image captions and a statistical relation called subsumption are used to generate term clusters which are organized hierarchically. Previously, the approach has been studied for document retrieval and results have shown that automatically generating hierarchies can help users with their search task. In this paper we present an implementation of concept hierarchies for image retrieval, together with preliminary ad-hoc evaluation. Although our approach requires more investigation, initial results from a prototype system are promising and would appear to provide a useful summary of the search results.

1. INTRODUCTION

One process that users must perform when information seeking is to examine and interpret the search results. In many Information Retrieval (IR) systems, results are ranked in order of relevance to the query. However, if many search results are returned it can be difficult for the user to examine them all. In addition, reliably providing an intuitive summary of the search results is an obvious benefit to any user of an IR system. Hearst [1] discusses various interface techniques for summarising results to make the document set more understandable to the user. These include: visualising the relationship of documents to the query, providing collection overviews and highlighting potential relationships between documents.

A variety of *clustering* techniques have been developed in IR to group documents into topically-coherent. This can help users to browse through the search results, obtain an overview of their main topics/themes and help to limit the number of documents searched or browsed in order to find relevant documents (i.e. limit exploration to only those clusters likely to contain relevant documents). Two common variations are: (1) to group documents by associated terms (i.e. a set of words or phrases define a cluster and membership is based on its containing a sufficient fraction of a cluster's terms), and (2) to assign documents to pre-defined thematic categories (manually or automatically). Scatter/Gather [3][4] and the Vivisimo[1] metasearch engine are an example of the former and Yahoo! Categories an example of the latter.

Organizing a set of documents automatically based upon a set of categories (or concepts) derived from the documents themselves is an obviously appealing goal for IR systems: it requires little or no manual intervention (e.g. deciding on thematic categories) and like unsupervised classification, depends on natural divisions in the data rather than pre-assigned categories (i.e. requiring no training data). In this paper we make use of such an approach for organizing search results called concept hierarchies [2][3]. This simple method of automatically associating terms extracted from a document set has been successfully used to help users searching and browsing for documents [4]. In this simple method, words and noun phrases (called concepts) are extracted from passages of the top *n* documents and organized hierarchically based on document frequency and a statistical relation called subsumption.

Given the simplicity of this method and its success for document retrieval, in this paper we apply concept hierarchies to textual metadata associated with images for image retrieval. There are many instances of when images are associated with some kind of text semantically related to the image (i.e. metadata or captions). For example, collections such as historic or stock-photographic archives, medical databases, art/history collections, personal photographs (e.g. Flickr.com) and the Web (e.g. Yahoo! Images). Retrieval from these collections is typically supported by text-based searching which has shown to be an effective method of searching images [5]. To enhance such systems, various approaches have been explored to organize search results based on either textual and visual features (or a combination of both). A summary of related work is provided in section 2. In practice, given the proliferation of textual metadata, investigating methods to exploit this text (e.g. for organizing results) is beneficial.

The paper is ordered as follows: in section 3 we describe how we use concept hierarchies as a method for presenting image search results by displaying extracted concepts within a hierarchical structure. This approach is fully automatic, generated from the search results, requiring no manually generated and assigned categories but simply associated text. We describe a working prototype system and present some preliminary results from analyzing the hierarchies. We discuss further plans for evaluation and uses for this clustering technique.

2. RELATED WORK

For image retrieval, clustering methods have been used to organize search results by grouping the top *n* ranked images into similar and dissimilar classes. Typically this is based on visual similarity and the cluster closest to the query or a representative

---

[1] http://vivisimo.com

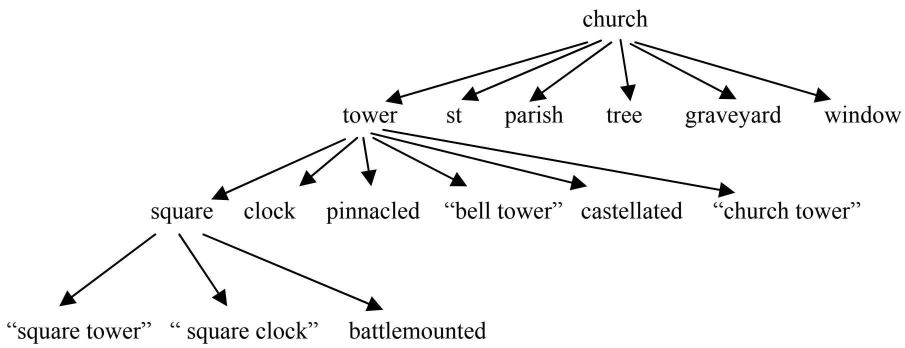

Figure 1: Example fragment of a concept hierarchy for the query “church” based on images in the St Andrews collection.

image from each cluster can then be used to present the user with very different images enabling more effective user feedback. For example Park et al. [6] take the top 120 images and cluster these using hierarchical agglomerative clustering methods (HACM). Clusters are then ranked based on the distance of the cluster from the query. The effect is to group together visually similar images in the results.

Other approaches have combined both visual and textual information to cluster sets of images into multiple topics. For example, Cai et al. [7] use visual, textual and link information to cluster Web image search results into different types of semantic clusters. Barnard and Forsyth [8] organize image collections using a statistical model which integrates semantic information provided by associated text and visual features provided by image features. During a training phase, they train a generative hierarchical model to learn semantic relationships between low-level visual features and words. The resulting hierarchical model associates segments of an image (known as *blobs*) with words and clusters these into groups which can then be used to browse the image collection.

Approaches using only semantic information derived associated text have also been used to organize search results and to aid browsing. For example, Yee, et al. [9] describe Flamenco, a text-based image retrieval system in which users are able to drill-down results along conceptual dimensions provided by hierarchically faceted metadata. Categories are automatically derived from Wordnet synsets based on texts associated with the images, but assignment of those categories to the images is then manual. Finally, Rodden et al. [10] performed usability studies to determine whether organization by visual similarity is actually useful. Interestingly, their results suggest that images organized by category/subject labels or were more understandable to users that those grouped by visual features.

3. BUILDING CONCEPT HIERARHCIES

The approach of building a concept hierarchy proposed by Sanderson and Croft [2] aims to automatically produce, from a set of documents, a concept hierarchy similar to manually created hierarchies such as the Yahoo! categories. The main difference being that concepts are in fact words and phrases (referred to as *terms*) found within the given set of documents and not categories defined manually (see, e.g. Figure 1). In their method of building concept hierarchies, word and noun phrases (called concepts) are extracted from retrieved documents and used to generate a hierarchy. Concepts are associated based on the set of documents indexed by the two concepts: the more documents two terms share, the more similar they are. However, concept hierarchies go beyond simple grouping of terms by discovering whether concepts are also related hierarchically. Document frequency and a statistical relation called subsumption is used to generate a hierarchy by detecting whether a parent term refers to a related, but more general concept than its children (i.e. whether the parent’s concept subsumed the child’s). Using document frequency (DF) to determine the semantic specificity of concepts is commonly used for weighting terms in IR based on Inverse Document Frequency (IDF).

With subsumption, concept C_i is said to subsume concept C_j when a set of documents in which C_j occurs is a subset of the documents in which C_i occurs. Or more formally, when the following conditions are held: $P(C_j|C_i) \geq 0.8$ and $P(C_i|C_j) < 1$. The assumption is that C_i is likely to be more general than C_j because, first, the former appears more frequently than the latter [13], and second, the former subsumes a large part of C_j’s document set. Also they are likely to be related since they co-occur frequently within documents. The results can be visualised using cascading menus where more general terms are placed at a higher level followed by related but more specific terms (Figure 2).

Sanderson and Croft analysed a random sample of parent-child relations and found that approximately 50% of the subsumption relationships within the concept hierarchies were of interest and that the parent was judged to be more general than the child. In particular, 49% of children were judged to reflect an aspect of the parent (a holonymic relation), e.g. actor is an aspect (or part) of a movie, 23% judged as a type of the parent (a hypernymic relation), e.g. a poodle is a type of dog, 8% judged to be the same as the parent, 1% as opposite to the parent, and 19% to be an unknown relation. We discuss relations commonly found using image captions in section 5. In summary, to generate a concept hierarchy for image browsing, the following steps are followed after an initial retrieval:

1. Extract concepts (words and noun phrases) from up to the top *n* image captions.

Figure 2: Example fragment from generated menu for the query "church"

2. Compare each concept with every other concept and test for subsumption relationships.
3. Order concepts hierarchically based on DF scores (general to specific) and subsumption relation (concepts with no parent – no other concept subsumes - are top-level concepts).
4. Randomly select an image from the cluster to represent the cluster visually and create the menu.

For our image retrieval prototype, we used a version of the CiQuest system created to investigate user interaction with a standard textual document collection [8]. The system uses a probabilistic retrieval model based on the BM25 weighting function [11] to perform initial retrieval. A DHTML menu is generated dynamically representing the concept hierarchy, enabling users to interact with and browse the search results (Figure 2). The number in parenthesis is document frequency.. A number of parameters can be adjusted in the prototype including:

1. *menu_depth*: maximum depth of menu;
2. *menu_height*: maximum height of menu;
3. *top_n*: number of documents to extract concepts from.

4. TEST IMAGE COLLECTION

The dataset used consisted 28,133 historic photographs from the library at St Andrews University[2]. All images are accompanied by a caption consisting of 8 distinct fields (short title, long title, description, location, date, photographer, notes and topic categories) which can be used individually or collectively to facilitate image retrieval. The 28,133 captions consist of 44,085 terms and 1,348,474 word occurrences; the maximum caption length is 316 words, but on average 48 words in length. All captions are written in British English and contain colloquial expressions and historical terms. Approximately 81% of captions contain text in all fields, the rest generally without the description field. In most cases the image description is a grammatical sentence of around 15 words. The majority of images (82%) are black and white, although colour images are

[2] http://specialcollections.st-and.ac.uk/

also present. The dataset has been used for previous image retrieval experiments, the most notable being the ImageCLEF evaluation[3] campaign for cross-language image retrieval [12].

5. PRELIMINARY EVALUATION

5.1 Experiment

The authors performed a preliminary ad-hoc evaluation and analysis of the concept hierarchies using images from the St Andrews photographic collection. We aimed to investigate the usefulness (and quality) of the hierarchies generated in a context different to standard document retrieval. Five example queries were selected from the ImageCLEF 2005 set of topics and used to retrieve a set of images: (1) Steam ships docked, (2) Stone viaducts, (3) Dog in sitting position, (4) Buildings covered in snow and (5) Fishermen in boat. Parent-child concepts were compared and the following relationships examined (based on Sanderson and Croft [2]):

- Child is *type-of* parent, e.g. a church > parish church, street > city street, ship > passenger ship;
- Child is *aspect-of* parent, e.g. church > window (part-of), masted steamer > SS Monoadock (instance-of);
- Child is the *same-as* the parent (synonymous) in the given context, e.g. hut > shed, train > engine;
- Child is *opposite-of* the parent, e.g. cat > dog;
- *Other*: another relation between parent-child potentially "interesting" to users.

Words and noun phrases are extracted from the top 1000 image captions. The *description* field of the caption is indexed and used for search and generating the concept hierarchies. We use only this field because it tends to describe the visual properties of an image (e.g. objects and relations). Search is performed using BM25[4] with default parameter settings. We experimented with setting the hierarchy depth and height and found a depth of 4 and height of 6 to provide the best results.

5.2 Results and observation

5.2.1 Analysis of parent-child relations

Table 1 shows the results of comparing, in total, 524 parent-child pairs for the 5 example topics given in section 5.1. On average, 68% of the pairs have some kind of "interesting" relation. We have found that many of the unhelpful pairs are due to unknown concepts (e.g. names of unfamiliar places of domain-specific words and phrases). Although 25% of the pairs exhibit the hierarchical relations of "aspect-of" and "type-of", there are many other interesting relations which could help users browse the results.

[3] http://ir.shef.ac.uk/imageclef/

[4] We also tried using a Boolean *anding* between query terms, but this resulted in too few results to generate a hierarchy from.

In general we notice that concepts are ordered from the general to more specific as expected using the subsumption relation, although we many further interesting relations exist other than just *aspect-of* and *type-of*. In general there appears to be two kinds of relations that exist between parent-child relations: (1) *Query-independent (Global)* - those which could be found in a global thesaurus such as Wordnet (e.g. dog is a type-of animal; door is a part-of car) and (2) *query-dependent* - those which identify relations between concepts specific to a query and/or domain (e.g. churches > Scotland).

Table 1: Relations between parent-child concept pairs

Qry	Pairs	%type -of	%aspect -of	Same	Oppo-site	%other
1	110	17%	10%	0%	0%	37%
2	104	20%	6%	1%	0%	53%
3	88	15%	11%	1%	0%	43%
4	135	10%	16%	0%	0%	40%
5	87	13%	9%	0%	0%	42%
Avg	***105***	***15%***	***10%***	***0.2%***	***0%***	***43%***

We found many examples where concepts are related, but their association is unobvious from the concept label alone. In particular, we have found many situations in which the parent-child concepts are not related lexically, but together describe visual properties of an image. For example, the parent-child pair "church > tree" would seem to be unrelated, but in fact the image contains pictures of both a church and trees. In this case, the ordering of the concepts is based purely on co-occurrence (i.e. there are more pictures where the term church is used than church and tree together) and is symmetric (i.e. it does not matter which order the concepts appear in). This is useful when browsing pictures that contain multiple concepts, but the ordering is unimportant. Because we use the description field of the captions, we might expect this to occur frequently because the words and phrases are being used in conjunction to describe the image. For example, in Figure 3 the associated text describes objects in the foreground and background (e.g. the concept pair: "snow-capped peak > heathery slopes").

Stately conifers to fore on heathery slopes above loch and around far banks; light cloud above snow-capped peaks beyond.

Figure 3: Example image and description field.

Based on our preliminary study of the concept hierarchies for this image collection, we found at least the following common types of relations between concept pairs:

- Child is *visually related* to parent, e.g. church > tree, people > paddling, stone > wall. In these cases both concepts can be seen in the image (AND relation);
- Child is *conceptually related* to the parent, e.g. cart > horses, dog > shepherd, train > station;
- Child is *description of* parent, e.g. building > tall building, suit > black suit, view > distant view;
- Child is *included in* parent hierarchy, e.g. tower > square > square tower (an error in the current system);

The success of the concept hierarchies can be seem to vary across different topics, but we have observed cases when the hierarchies are very effective at potentially minimizing the amount of browsing a user must perform. For example, for the query "buildings covered in snow" the top-level parent concepts include "snow" and "mountain". It is unlikely that pictures of buildings are found under "mountain" thereby limiting investigation to the menus under "snow".

6. DISCUSSION

Figures 2 and 4 show fragments of the menus generated for the single-word query "church", and the multi-word query "stone viaducts". In general the ordering of concepts appears to go from general to more specific, e.g. "church > tower > square" and "viaduct > railway > railway bridge". The top-level concepts include the query terms and we observe that in the case of multi-word queries, depending on co-occurrences between document terms, typically each query term is represented by a separate top-level category. From the example menu fragments shown in Figures 2 and 4, we make the following observations:

- In many cases, the images provide useful additional contextual information to clarify the concepts. For example, it may not be clear to a user what an "estuary" is, but the picture displayed for this concept may help.
- We have noticed that many of the concepts include geographical locations (collection-specific information). For example, in Figure 3 concepts such as Cumnock and Lochearnhead are locations in Scotland. These are likely to be useful only if the user is familiar with the collection. Providing some kind of indication as to the type of concept could be beneficial.
- The images displayed may help to disambiguate potentially ambiguous concepts, e.g. for the concept "Fall" under "Dog", the image shows a waterfall rather than an animal. The concept indeed refers to a waterfall called "Dog Fall". Other examples include "Man > Spring" where "spring" refers to a source of water.
- We have observed that synonymous concepts often appear within the same level of the hierarchy and should probably be merged. For example, "masted fishing *boat*" and "masted fishing *vessel*" appear in the same menu level or column.
- The selection of "representative" images for a concept is problematic and often confusing (e.g. the pictures of "stone" and "hill" in Figure 3). This highlights the limitation of selecting the images randomly. This is especially observant when the concept contains images which are visually dissimilar, e.g. "view".
- Concepts are varied than typical subject categories (like Yahoo!). Being automatically generated, the quality of the concepts varies, but sometimes hierarchies are generated which are highly descriptive of a set of images, e.g. for Figure 3, the following hierarchy is generated "Peak > snow-capped peaks > heathery slope". In Figure 4, the phrase "deep valley" is a child of the parent "Railway Viaduct" which in conjunction with the image is potentially useful when searching.

7. CONCLUSIONS

In this paper, we have showed how a simple term association method called concept hierarchies can be to automatically cluster co-occurring terms found in image captions and arrange terms hierarchically using a statistical relation called subsumption. We have developed a prototype system to demonstrate this approach for organising images based on concepts derived from image captions in the St Andrews historic photographic library. An initial investigation has shown that the relation between concept pairs is often useful and interesting when browsing a set of images. By adding an example image from the concept's cluster, context is added to the hierarchy helping to understand the concepts and their relations.

Concepts and relations generated with this automatic method vary widely from those typically found in global thesauri; to those which are more collection-specific. Although in general the concepts are ordered in a hierarchical manner by the subsumption relation, there are also many other interesting relations which are not "aspect-of" or "type-of" or in a hierarchical relationship. Based on our initial investigation, we feel this method for organising image retrieval results (when images are accompanied by texts) is a promising technique and warrants further investigation.

8. FUTURE WORK

Much of our planned further work will be evaluation, but there are a few improvements we would like to make to the existing system. The first is to remove erroneous relations such as "church > tower > church tower", the second is to highlight geographical locations to users and the final improvement would be to group synonymous concepts within the same hierarchy level. Given the success of Flamenco, we would also like to experiment with assigning words more abstract concepts from Wordnet and then clustering these. It might help to remove terms or concepts which are too specific and unlikely to help users in general.

Figure 4: Example fragment from generated menu for the query "stone viaducts"

So far we have only carried out preliminary evaluation of the concept hierarchies for image retrieval and there are many interesting avenues we would like to explore. For example, we plan to carry out a user study in a manner similar to Joho et al. [4] where the hierarchies are evaluated in a practical retrieval setting. We also plan to contrast the success of hierarchies for retrieval based on displaying text-only, image-only and then text and image. We would like to investigate further their usefulness individually and collectively for generating a visual concept hierarchy. Another planned experiment will use the ImageCLEF test collection to compare where relevant images appear in a ranked list and the hierarchy structure. This will enable us to postulate the amount of searching through the menus a user would be required to do in order to find relevant images. Finally a number of question we would like to address include the following:

- How would users actually like images in the St Andrews collection organised?
- Is the representative hierarchy image suitable – how is this selected?
- Does the selected hierarchy keyword match the displayed image?
- Do the images create a hierarchical appearance – are parents more general than children?
- Do images in the menus represent useful relations?
- Does the image represent the concept well?
- Is chosen image representative of the cluster of images?

9. REFERENCES

[1] Hearst, M. (1999). "User Interfaces and Visualization". In: Baeza-Yates, R. & Ribeiro-Neto, B. (eds.), *Modern Information Retrieval*, pp. 257-323. New York: ACM Press.

[2] Sanderson, M. and Croft, B. (1999) "Deriving concept hierarchies from text" In: *Proceedings of the 22nd ACM Conference of the Special Interest Group in Information Retrieval*, pp. 206-213.

[3] Sanderson, M. and Lawrie, D. (2000) "Building, Testing, and Applying Concept Hierarchies" In: W. Bruce Croft,

(ed.), *Advances in Information Retrieval: Recent Research from the CIIR*, Kluwer Academic Publishers, pp. 235-266.

[4] Joho, H., Sanderson, M., and Beaulieu, M. (2004) "A Study of User Interaction with a Concept-based Interactive Query Expansion Support Tool". In: McDonald, S. & Tait, J. (eds), *Advances in Information Retrieval, 26th European Conference on Information Retrieval*, pp. 42-56.

[5] Markkula, M. and Sormunen, E. (2000) "End-use searching challenges indexing practices in the digital newspaper photo archive", Information Retrieval, **1**, pp. 259-285.

[6] Park, G., Baek, Y., and Lee, H-K. (2005) "Re-ranking algorithm using post-retrieval clustering for content-based image retrieval", *Information Processing and Management*, **41**(2), pp. 177-194.

[7] Cai, D., He, Xiaofei., Li, Zhiwei., Ma, W-Y., and Wei, J-R. (2004) "Hierarchical clustering of WWW image search results using visual, textual and link information". In: *Proceedings of the 12th annual ACM international conference on Multimedia*, pp. 952-959.

[8] Bernard, K. and Forsyth, D. (2001) "Learning the Semantics of Words and Pictures". In: *Proceedings of the Intentional Conference on Computer Vision*, vol 2, pp. 408-415.

[9] Yee, K-P., Swearingen, K., and Hearst, M. (2003) "Faceted metadata for image search and browsing". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 401-408.

[10] Rodden, K., Basalaj, W., Sinclair, D., and Wood, K. (2001) "Does Organisation by Similarity Assist Image Browsing?", In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 190-197.

[11] Robertson, S.E., Walker, S., Beaulieu, M.M., Gatford, M. & Payne, A. (1995). "Okapi at TREC-4". In: Harman, D.K. (ed.), *NIST Special Publication 500-236: The Fourth Text REtrieval Conference (TREC-4)*, Gaithersburg, MD. pp. 73-97.

[12] Clough, P., Mueller, H. and Sanderson, M. (2005), "The CLEF 2004 Cross Language Image Retrieval Track", In: Peters, C., Clough, P., Gonzalo, J., Jones, G., Kluck, M. and Magnini, B. (Eds.) *Multilingual Information Access for Text, Speech and Images: Results of the Fifth CLEF Evaluation Campaign*, Lecture Notes in Computer Science, Springer, to appear.

[13] Sparck Jones, K. (1972). "A Statistical Interpretation of Term Specificity and its Application in Retrieval". *Journal of Documentation*, 28 (1), 11-21.