Image Based Gisting in CLIR

Mark Sanderson
Department of Information Studies,
University of Sheffield,
Regent Court, 211 Portobello St,
Sheffield, S1 4DP, UK
+44 (0) 114 22 22648
m.sanderson@shef.ac.uk

Robert Pasley
Department of Information Studies,
University of Sheffield,
Regent Court, 211 Portobello St,
Sheffield, S1 4DP, UK
+44 (0) 114 2257683
rpasley@blueyonder.co.uk

ABSTRACT
In this paper, we describe research which could lead to a novel approach to gathering an overview of a document in a foreign language. The research explores how much of the meaning of a document could be represented using images by researching the ability of subjects to derive the search term that might have been used to return a set of images from an image library. The Google image search engine was used to retrieve the images for this experiment, which uses English throughout. The results were analysed with respect to a previous paper [1] exploring ability to recognise concrete objects in hierarchies. It was found that there is a tendency to use one particular level of categorization.

Categories and Subject Descriptors
H.3.3 Information Search and Retrieval

Information filtering, Query formulation, Relevance feedback, Retrieval models.

General Terms
Languages

Keywords
cross language information retrieval, gisting, images, basic level.

1. INTRODUCTION
This paper describes research into the feasibility of using images to allow understanding of languages unknown to a user whilst searching in electronic documents. A survey was carried out to gather information on subjects' abilities in this respect. Search terms were defined and images found using an online image library (Google Image). An established theory from cognitive physiology is used to group the search terms and responses depending on their level in a hierarchical structure. The changes of level between search term and response were examined, as was user performance in different levels.

2. METHODOLOGY
A paper by Rosch [1] states that people categorise concrete objects with reference to a basic level which is;

- "The most inclusive category for which a concrete image of the category as a whole can be formed
- The first categorizations made during perception of the environment
- The earliest categories sorted and earliest named by children
- The most codable, most coded and most necessary in language" [1]

The experiment was designed in the form of a web based survey; emails were sent to 283 staff and students of the Department of Information Studies at the University of Sheffield. The questionnaires consisted of 25 sets of 8 images, one using 6 optional multiple choice questions and one a free text box. Other data was collected about the subjects and their level of familiarity with information searching.

The search terms were derived from two sources. Firstly a paper by Rosch [1] describing the basic level of categorization, secondly the researcher created a set of search terms following the principles of that paper. The search terms were selected to represent the following groups; abstract, concrete, ambiguous and noun + verb. Only concrete search terms were available from the Rosch [1] paper.

These search terms were entered into the Google Image search engine in order to find the images for the experiment. English was used for the image searching and the questionnaire. The top 8 images without text as part of the image were used for the questionnaire.

3. RESULTS

3.1 Response
Of those surveyed, 18% responded. The respondents felt themselves to be comfortable with information searching; on a scale of 1 to 5 there was a mean of 3.84 comfort level.

The responses to the questionnaire were categorized into levels with respect to the basic level, sub-ordinate level and super-ordinate level. The results were analysed to see whether the results were better in different search term levels.

It was found that 81% of the responses were correct in the basic level, 21% in the sub-ordinate level and 21% for super-ordinate search queries. Overall 37% of answers were correct. These statistics are based on exact and very close answers; for example clothing as a query had answers of Clothes, Clothing and Clothings; these were accepted. Shirts, T Shorts and Tops were not accepted.

SIGIR'04, July 25–29, 2004, Sheffield, South Yorkshire, UK.
ACM 1-58113-881-4/04/0007.

3.2 Analysis

It was found that when the search term was in the super-ordinate level the results tended to be in the super-ordinate level. Overall 41% of the results have been judged to be in the super-ordinate category. With 31% in basic and 14% in sub-ordinate there also appears to be a tendency for the basic-level to be used in the results.

For the sub-ordinate queries, responses were mainly in the basic level category. The super-ordinate levels are hardly used (4%). The sub-ordinate level has 17% of the results, but this is much less than the basic level at 72%.

Using basic-level search queries resulted in a very high percentage of basic-level answers (85%). There are no sub-ordinate level results and only 8% at super-ordinate level.

The searches from the super-ordinate level show the highest levels of answers judged as "non-taxonomical" or "not sure". These have combined percentages for the super-ordinate search terms of 15% as opposed to 7% for basic level and 8% for sub-ordinate.

Most of the results are within the basic level. For sub-ordinate searches this means the results are generalisations of the search queries. For basic level there is no change of level. For the super-ordinate search level many responses are also super-ordinate level. However, the super-ordinate level also returns a high percentage of basic level results meaning that that group was made more specific.

Analysing responses and search queries with regard to levels it was found that for sub-ordinate level search queries 17% of the results are sub-ordinate, for super-ordinate 41% of responses are super-ordinate, the basic level show the greatest consistency with basic-level searches returning 85% basic level responses.

4. ANALYSIS

Chandler [2] explains that when dealing with systems of signs understanding requires interpretation. The medium of transmission of information interferes with the meaning to some extent due to the effectiveness of the medium at representing certain types or aspects of information. Systems of comparison are commonly used to "make the unfamiliar more familiar" [2]. Since they all involve comparison with culturally negotiated exemplars they are part of the culture in which they develop. Standard metaphors, for example, are generally speaking language dependant.

For the above reasons it can be argued that images can be used to represent concepts in the same way as text. The question becomes one of whether sufficient exemplars exist that can be used as visual tropes (or symbols).

Some authors view images as a medium significantly different to text. Ernst Gombrich quoted in Chandler [2] states that 'pictures cannot assert', suggesting that a large number of texts could not be represented wholly through images. Barnard et al [3] describe a system that takes advantage of the convergence of the fields of image recognition and information retrieval indexing. They describe an image region labeling system which has its vocabulary restricted by the text associated with the image. Within certain bounds images can certainly describe, and this may be useful in the multi-lingual gisting of documents.

Some search terms were very well anticipated, whilst others could not be guessed from the images shown. Some were right but not absolutely right, some were generalised by all respondents. Note that some of these were proper nouns and given images of the specific item a more descriptive, higher level word was chosen. Rosch [1] describes the concept of a prototypical object within a category. These are more representative of the category as a whole. An example would be that "Robin" is more typical of the category "Bird" than "Penguin". It may be that some of the proper nouns are highly prototypical in their category and therefore evoked the category rather than the specific item. It may also be that the distinguishing factors are not clear to the majority of respondents. The sub-ordinate searches showed a similar generalising effect. One super-ordinate search was also generalised in the results. Although some questions did have more specific answers none had significantly more specialised answers than expected.

5. DISSCUSSION

The experiment revealed that there is a tendency to answer with basic level responses in search queries in all three levels. This means that sub-ordinate search queries are badly represented with a tendency for the image representation to be a generalisation of the search query meaning. The representation at super-ordinate level also tends to lose meaning, because respondents focus on few specific images. However the loss of meaning is not so pronounced as the sub-ordinate level. Search queries in the basic level perform best. This is especially true if they represent something that could be termed a concrete object.

6. CONCLUSION

It seems that in order to represent documents using images it is important to find out the level of the words as described by Rosch [1] which would be used as search terms. The research also suggested that the more concrete a search term the better chance it stood of being represented using images. Future research might lead to methods to identify these search terms in order to automatically decide how to represent them.

7. REFERENCES

[1] Rosch, E. Mervis, C., Gray, W, Johnson, D and Boyes-Braem, P. (1976). "Basic Objects in Natural Categories." Cognitive Psychology, 8, 382-439.

[2] Chandler, D. (2002). Semiotics: The Basics. London: Routledge.

[3] Barnard, K., Johnson M. and Forsyth, D. (2003). "Words Sense Disambiguation with Pictures". In: Workshop on Learning Word Meaning from Non-Linguistic Data (2003) Held in Conjunction with The Human Language Technology Conference. 27 May – 1 June, 2003. Edmonton, Canada.