

# Exploratory Search: Image Retrieval without Deep Semantics

John I. Tait

School of Computing and Technology, University of Sunderland, St. Peter's  
Campus, Sunderland SR6 0DD, UK

[John.Tait@sunderland.ac.uk](mailto:John.Tait@sunderland.ac.uk)

**Abstract.** This paper relates semantics as it is used in linguistics and natural language processing to the operational requirements of image retrieval systems. This is done in the context of a model of exploratory search and image annotation or indexing. The paper concludes this operational context requires the use of a restricted form of semantics compared with the usual one from linguistics or natural language processing, focussing on words rather sentences.

## 1 Introduction

Semantics is a notoriously difficult topic. Within Linguistics, on which most work in computerised Natural Language Processing (NLP) rests, it is often all but impossible to draw clear distinctions between syntax, semantic, and pragmatics let alone define the content of semantics alone.

However, this is not a doctrine of despair: in this paper I want to show how within a given operational context the general problems of semantics can be avoided, or at least restricted to the extent they are solvable in the medium term.

More generally I want to look at the task of retrieving still images from databases. I will attempt to show how restricting ourselves to this task allows us to avoid such general and difficult questions as “what does this image mean?” which I believe are the root of much of the difficulty with the notion of semantics.

By both recognising that the real issues relate to operations involving humans (not abstract ones involving computers alone) and recognising that at some point this allows a human understanding (or subjective semantics if you will) to be brought to bear on the operational task at hand, one can focus on much more tractable questions like “Is this image relevant to the ends of this person at this time?”. This can be explored in a framework of exploratory search

---

Please use the following format when citing this chapter:

Tait, John, 2006, in IFIP International Federation for Information Processing, Volume 204, Artificial Intelligence Applications and Innovations, eds. Maglogiannis, I., Karpouzis, K., Bramer, M., (Boston: Springer), pp. 566–574

The paper begins with a brief overview of the use of the term semantics in Linguistics and NLP. The idea of exploratory search is then introduced, developed using a model and related to recent developments in image annotation. The paper concludes with an attempt to bring together the implicit notions of *semantics* inherent in exploratory search and image annotation with the linguistic notion of semantics previously introduced.

## 2. Semantics in Linguistics and NLP

Semantics is of course the study of meaning.

James Allen in his classic text book [1] identified seven kinds of knowledge relevant to natural language understanding (including speech): phonetic and phonological; morphological; syntactic; semantic; pragmatic; discourse; and world knowledge. Syntactic knowledge:

“concerns how words can be put together to form correct sentences and determines what structural role each word plays in the sentence and what phrases are subparts of what other phrases” (p10)

whereas semantic knowledge:

“concerns what words mean and how these meanings combine in sentences to form sentence meanings. This is the study of context-independent meaning – the meaning a sentence has regardless of the context in which it is used.” (p10)

Note the emphasis Allen places on word meaning and the way they combine presumably constrained and driven through the syntactic constructions in which they appear.

Even within this constrained notion of (linguistic) semantics there are further complications. Leech [17] within an essentially compatible framework identifies (oddly) seven types of meaning, in essentially three groups: Conceptual, Associative and Thematic. Lyons [18] effectively identifies ten types of meaning, through the rather recursive device of looking at the meanings of the noun “meaning” and the verb “to mean”.

Allen, and Leech (the latter less clearly) are committed to the idea that words have distinct senses or meaning<sup>1</sup>. But even this is controversial: see Kilgarriff [16] for a discussion. The English word “bank” with its two homonyms (“financial institution” and “river bank”) is well known. More difficult examples are not hard to find. Consider the word “drug”. Are sentences in which it is used interchangeably with “narcotic” using a different sense from those in which it is used interchangeably with “medicine”?

The reason for this short discussion of semantics as the term is understood by the natural language understanding and linguistics communities is to contrast this with the way it used in the multi-media and image retrieval communities. Jörgensen [13] (p167ff) and more recently, for example, Koskela and Laaksonen [15] and Heesch and Rüger [11] use the term with a much narrower meaning: semantic in the context

<sup>1</sup> Lyons position is somewhat more sophisticated considering, for example the notion of “linguistic fields”.

of image retrieval is understood to be confined to assigning conceptual labels from some sort of fixed vocabulary to the whole or parts of a region, possibly with some sort of spatial relations between regions identified.

This is much more restricted notion of semantics than is commonly employed in natural language processing. There is nothing wrong with it (especially given the state-of-the art in multimedia retrieval and how far multimedia retrieval has come in the past ten years or so): but a failure to recognise the differences in the ways the term is being used is likely to lead to confusion.

It might be better to recognise that what is adopted in the multimedia field is an *operational semantics*, in which meaning is defined in terms of specific operations and tasks, with no claim to deal with the general meaning of multimedia data. Indeed many would claim that adopting restrictions in terms of task and domain is precisely where progress has been made in both natural language processing and the semantic web.

### 3. Exploratory Search

The real focus of the paper is a new way of looking at the process of multimedia retrieval: **exploratory search**.

Exploratory search is a notion which applies to all forms of information seeking, but is especially applicable to multimedia retrieval. It focuses on situations where the searcher has an ill specified information need. It cuts across the commonly used searching characterisation of *information discovery* [14] versus *previously seen information retrieval* (e.g. Dumais et al [9]) in which the primary characterisation is whether the search system is mainly intended to support the retrieval of items previously not seen by the searcher (information discovery) versus those previously seen but whose virtually whereabouts are presumably currently unknown.

Most current or proposed exploratory search systems have some sort of topological metaphor underlying both their indexing structures and their interfaces. A typical system might involve some sort of categorisation or clustering step which is used to identify key examples or summaries, an associated similarity space or spaces which allow the topological space to be projected onto a two or two-an-a-half dimensional display space and some form of analytical query to allow users to “parachute” in to suitable areas of the topological space without considering the whole of it [6,28].

Some systems focus more on guided interaction and the process of negotiating the search, but the common theme is the metaphor of exploring an underlying space.

I don't want to claim that exploratory search is anything new or radical: indeed many of the ideas can be traced back to Belkin and others [5] and the notion of Anomalous States of Knowledge, if not before.

Neither do I want to pretend the term is my own<sup>2</sup>.

Digital Still Image Retrieval provides an interesting challenge for emerging exploratory image search systems. This is for three reasons. First, the content of

<sup>2</sup> See <http://www.umiacs.umd.edu/~ryen/xsi/>

images is rich, multi-faceted and complex (an image is worth a thousand words). Second, user needs are highly subjective (even compared to text) and difficult to define (browsing for the pleasure of browsing, my grandchildren give me joy, whereas yours are merely children). Third it is possible to attend to and take in many images more or less simultaneously, in a way which is not possible with text and music for example.

This has is leading to the emergence of a new field which I think is best called Semantic Content Based Image Retrieval (SCBIR) in contrast to Content Based Image Retrieval based on low-level features [23].

Before moving on to a more detailed discussion of SCBIR and Exploratory Search I want to look more carefully at the term semantics and way it is understood in natural language processing and linguistics.

#### 4. A Model for Exploratory Image Search

As with most interactive information retrieval systems the first step is to split the retrieval into two: an indexing process which can take place (slowly) offline and an interactive retrieval process which of necessity must operate rapidly and be focussed on effective and “comfortable” use by real people undertaking searches.

The indexing process then takes on a subsidiary role in which data is selected and organised to allow rapid and effective organisation of the interactive retrieval engine. Although the indexing process is in fact technically more demanding, since it is logically subsidiary I want to consider first the exploratory search process.

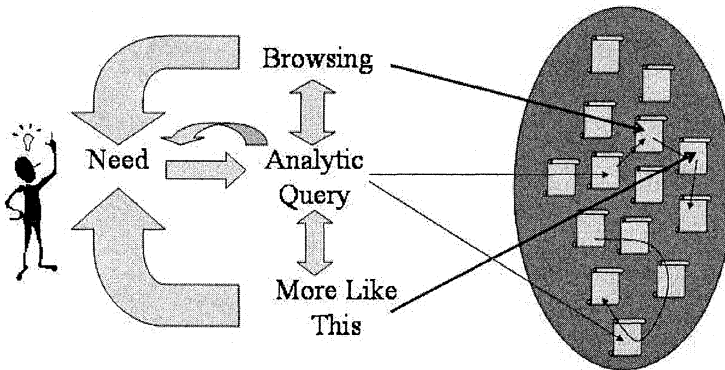


Fig. 1. A Model for Exploratory Image Retrieval

Figure 1 presents a model for exploratory image retrieval. In it a searcher first comes to the system with some sort of need or requirement for an image. This they express as an analytic query, for example as a list of key words.

This will produce an initial retrieved set of images. These may well change the searcher's idea of their need. The retrieved set may also act as a starting point for other forms of retrieval: for example requests for more images in some sense similar to one of the retrieved set: or less directed browsing. The images seen during these subsequent search operations may in themselves change or refine the searcher's idea of their need. Eventually, presumably, the searcher will find a satisfactory image set of images or give up.

Now I make no great claims for this model. It is very similar in outline to ones presented, for example, in Belew (2000) Chapter 1. Compared to Belew, following Belkin, Oddy and Brooks (1982) there is perhaps, a little more emphasis on the impact of the results of retrieval on the searcher's perception of their own need, but this is a matter of detail.

More to the point is the reliance of the model, on the one hand, on comprehensible notions of semantic similarity, and on the other on an effective and efficient means of indexing the images to support analytic query.

Indexing to support analytic query is primarily a process of assigning appropriate key words to images. I will return to this point later.

Semantic similarity is needed to support both browsing and relevance feedback or more-like-this access. Human notions of image similarity are complex (see [13], Chapter 2 for a review). Browsing often also requires some notion of directionality or scent [8] so that the user can browse successive image sets which have more or less of some property.

Current systems operate with similarity computed using low-level features: in particular colour (see for example [19], [20]), and this is an impoverished notion of image semantics.

Browsing and search are essential features of exploratory search systems, but despite my title I don't want to dwell on them here: rather I want to move to look at a more underlying problem – the problem of associating key words with images.

## **5. SCBIR Indexing for Analytic Query: Automatic Image Annotation**

Within the model of Section 4 the first step in exploratory search is for the searcher to formulate their need into an analytic key word query.

If we are to build exploratory search systems which are capable of operating on a very large scale, we require a rapid, automatic, means of annotating images with key words.

This is a daunting problem: the subjectivity of the understanding of images, their complexity and the semantic gap caused by the differences between the low-level features of images used in computerised processing of digital images and human understanding, all contribute to the significant progress which has been made with this problem in recent years.

The semantic gap also occurs in other forms of non-textual image retrieval (like video and music) but the problems it poses and the ways to overcome them are distinctive in still image searching partially because of some technical problems

posed by still images (meaningful segmentation of still images is especially hard) but equally, as noted above, still image perception has some special properties which may be exploited in the interface.

In particular it appears possible for searchers to rapidly or simultaneously scan very large numbers of still images in result sets. This contrasts with text, where, seemingly fuller attention needs to be applied to results viewed, and consequently the useful size of result sets presented. Video appear to present related problems, with distraction by non-relevant results being a greater problem.

However, it must be pointed out there is a need for much more extensive ecologically valid, human centred studies in this area before firm conclusions can be drawn. Markkula and Sormunen [21], for example, report their searchers frequently adopted what they perceived to be the best or easiest search strategy. It is unclear whether their perceptions were in fact correct.

However, there appears to have been real progress with bridging the semantic gap in the past few years. The successes have in common the use of various forms of supervised and semi-supervised machine learning. For clarity, in supervised machine learning the learning system builds a model on a hand analysed set of data called the training set. In this case this will be a set of images which have been manually annotated. Semi-supervised systems will either have some form of manual assessment or correction on additional unseen data or combine the use of entirely automatic, unsupervised learning with supervised learning.

The most successful of these early attempts employed models from machine translation, [2,3] although more recently there have been suggestions the relationship between the task and Machine Translation is more indirect [27]. Other forms of machine learning and adaptive computing have also proved successful [7, 10, 12, 24, 25, 26].

There are three main observations to be made about these pieces of work.

First, they generally focus on the assignment of unordered sets of key words to the whole of an image, since this is really what is required for indexing. Where they do not, they are not reliant on accurate, meaningful segmentation of the image. They have models which combine both the relationship between areas of the image and words and the corrections for the likely errors in those relationships.

Second, all have vocabularies which are much too small for practical applications. Current systems have vocabularies of a few hundred terms in contrast to the 20,000 plus terms in the Art and Architecture Thesaurus [13], for example.

Third, for the foreseeable future they are relatively errorful, in the sense that they will often assign words to images which would not seem relevant to a human being. However, in the context of exploratory search this is not a real problem. The human searcher can overcome the errors made by the indexing systems by using the browsing and relevance feedback mechanisms, provided of course, those errors are not so severe and numerous that the searcher see no relevant images in the early phases of their search.

## 6. The Relationship between Linguistic Semantics and Image Semantic for Exploratory Search

In Section 2 I overviewed the use of the word “semantics” as it is used in Natural Language Processing and Image Retrieval. I also introduced the term *operational semantics* to describe a situation in which no claim is made about the generality of the semantics without a task.

I now want to weave together the two threads of this paper: exploratory search and semantics.

The first observation I want to make is that there is a critical element in Allen’s definition of semantics discussed in Section 2 which is missing from all the subsequent discussion of exploratory search and image retrieval. Allen’s definition focuses on sentence meaning, rather than the meaning of words, which has been the focus in the subsequent discussion.

One might argue that this focus on the sentence is a weakness of most conventional NLP and linguistic work. However most work in NLP takes the sentence as the important unit of meaning.

However, operationally, in the context of exploratory search, few searchers wish to express their needs through well formed sentences.

Conversely, considering the image annotation or indexing process, there seems to be little utility in annotating images with long sentences rather than a bag of key words. (I accept it is plausible some phrases might be useful: e..g “ducks on water”.) Indeed annotating images with sentences opens up a real Pandora’s box: which are the important concepts/words to include in the sentence (depends on the context of use): if we use several sentences when to stop; and so on. These matters are only on the research agenda in the longest term.

My purpose here is to point out that within Natural Language Processing or Linguistics a focus on the meaning of words seems a very narrow use of the term semantics, Yet in moving to a specific operational task: like exploratory search; in order give an attainable scope and definition to our enterprises in multimedia we may need to focus on this narrow use at least in the short term.

## 7. Conclusions

In this paper I have attempted to do three things.

First I have attempted to overview the way the term “semantics” is used in Natural Language Processing.

Second, I introduced the idea of exploratory search and model for its use in image retrieval, including the use of automatic image annotation.

Third I attempted to relate the notion of semantics as it is used in NLP and Linguistics to the operational context of exploratory image search.

Semantics as the term is used in NLP and Linguistics implies a very deep and complex notion of meaning. By focussing on a specific, practical operational context (like exploratory search) we can avoid the possibly intractable problems posed for multimedia processing by these deep semantics.

## Acknowledgements

I would like to thank Peter Enser for originally introducing me to Corinne Jörgensen's excellent and comprehensive book. I would also like to thank Ryen White for introducing me to the idea of Exploratory Search and for making very useful comments on an earlier draft of this paper.

## References

1. Allen, J. *Natural Language Understanding* (2nd Ed.). Benjamin/Cummings, Redwood City, CA, USA. 1995.
2. Barnard, K. P. Duygulu and D. Forsyth, "Recognition as Translating Images into Text" *Internet Imaging IV, Electronic Imaging 2003* (Invited paper). IS&T/SPIE 2003.
3. Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D., and Jordan, M.I. (2003) Matching words and pictures. *Journal of Machine Learning Research*, vol. 3, pp. 1107-1135.
4. Belew, R.K. *Finding Out About*. Cambridge University Press, 2000.
5. Belkin, N.J., R.N. Oddy and H.M. Brooks. *ASK for Information Retrieval: Part I: Background and Theory*. *Journal of Documentation* 38(2), 1982. Reproduced in *Reading in Information Retrieval*. K. Sparck Jones and P. Willett (Eds). Morgan Kaufman, San Francisco, CA, USA. 1997.
6. Cai, D. He, X, Z. Li, Ma, W.-Y. and J.-R. Wen "Hierarchical Clustering of WWW Image Search Results Using Visual, textual and Link Information" *Proceeding of the 12th Annual ACM International Conference on Multimedia (MM '04)*, ACM Press, 2004. 952-959.
7. Carneiro, G., Vasconcelos N. "A Database Centric View of Semantic Image Annotation and Retrieval" *Proceeding of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005)*, Salvador, Brazil. ACM Press. 2005. 559-566.
8. Chi, E.H. P. Pirolli, J. Pitkow "The scent of a site: a system for analyzing and predicting information scent, usage, and usability of a Web site" *Proceedings of the SIGCHI conference on human factors in computing systems*, The Hague, The Netherlands, ACM Press 2000 161-168
9. Dumais, S. E. Cutrell, J.J. Cadiz, G. Jancke, R. Saran, D.C. Roberts "Stuff I've Seen: A System for Personal Information Retrieval and Re-use" . *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2003)*, Toronto, Canada. 2003.
10. Ghostal, A., P. Ircing, S. Khidanpu. "Hidden Markov Models for Automatic Annotation and Content-based Retrieval of Images and Video" *Proceeding of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005)*, Salvador, Brazil. ACM Press. 2005. 544-551.
11. Heesch, D. and S. Rüger, "Image Browsing: Semantic Analysis of NNk Networks" *Image and Video Retrieval, Proceeding of the 4th International Conference, CIVR 2005*. LNCS 3568 Springer 2005. 609-618.
12. Jeon, J. and R. Manmatha. (2004). "Using Maximum Entropy for Automatic Image Annotation". *3rd International Conference on Image and Video Retrieval (CIVR 2004)*, 24-32.



13. Jörgensen, C. *Image Retrieval: theory and Research*. Scarecrow Press, Oxford, UK. 2003.
14. Kearns, A and S.M. Smith "The Information Discovery Framework" Symposium on Designing Interactive Systems (DIS 2004), Cambridge, Mass. USA. 2004.
15. Koskella, M. and J. Laaksonen "Semantic Annotation of Image Groups with Self-organizing Maps", *Image and Video Retrieval*, Proceeding of the 4th International Conference, CIVR 2005.LNCS 3568 Springer 2005. 518-527.
16. Kilgarriff, A. I Don't Believe in Word Sense. *Computers and the Humanities* 31(2), March 1997. pp 91-113.
17. Leech, G. *Semantics*. Penguin. Harmondsworth, Middlesex, UK. 1974.
18. Lyons, J. *Semantics*. Cambridge University Press.1977
19. McDonald, S.. T.-Sh. Lai and J.I. Tait, "Evaluating a Content Based Image Retrieval System" Proceedings of the 24th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001), New Orleans, September 2001. W.B. Croft, D.J. Harper, D.H. Kraft, and J. Zobel (Eds).
20. McDonald S. and J.I. Tait "Search Strategies in Content-Based Image Retrieval" Proceedings of the 26th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003), Toronto, July, 2003. pp 80-87.
21. Markkula, M. and E. Sormunen "Searching for Photos: Journalists Practices in Pictorial IR". *The Challenge of Image Retrieval, a Workshop and Symposium on Pictorial IR*. Newcastle-upon-Tyne, UK. 1998.
22. Pulman, S.G. "Lexical Decomposition" in *Charting a new Course: Natural Language Processing and Information Retrieval – Essays in Honour of Karen Spärck Jones*. J.I. Tait (Ed.) Springer 2005.
23. Smeulders, A.W.M., Worring, M., Santini, S. Gupta, A., and Jain, R. (2000) Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380.
24. Srikanth, M., J. Varner, M. Bowden, D. Moldovan "Exploiting Ontologies for Automatic Image Annotation" Proceeding of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005), Salvador, Brazil. ACM Press. 2005. 552-558.
25. Tsai, C.F. K. McGarry and J.I. Tait "Automatic Metadata Annotation of Images via a Two-level Learning Framework" *ACM SIGIR Semantic Web Workshop*, Sheffield, July 2004. pp 32-42
26. Tsai, C.F., K. McGarry and J.I. Tait "Qualitative Evaluation of Automatic Assignment of Keywords to Images". *Information Processing and Management*, Volume 42 Issue 1, January, 2006. pp 136-154.
27. Virga, P., P. Duygulu "Systematic Evaluation of Machine Translation methods for Image and Video Annotation," *Image and Video Retrieval*, Proceeding of the 4th International Conference, CIVR 2005.LNCS 3568 Springer 2005.174-183.
28. Yee, K.-P., K. Swearingen, K. Li, M. Hearst. "Faceted Metadata for Image Search and Browsing" Proceedings of the SIGCHI conference on Human factors in computing systems, Ft. Lauderdale, Florida (CHI 2003). ACM Press. 2003 401-408.