City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

KNOWLEDGE ORGANISATION AND ITS ROLE IN MULTIMEDIA INFORMATION RETRIEVAL

Andrew MacFarlane
Department of Computer Science and Department of Library and Information Science, City University London, London EC1V 0HB, <andym@city.ac.uk>

ABSTRACT

Various kinds of knowledge organisation (such as thesauri) are routinely used to label or tag multimedia content such as images and music, to support information retrieval i.e. user search for such content. In this paper we outline why this is the case, in particular focusing on the semantic gap between content and concept based multimedia retrieval. We survey some indexing vocabularies used for multimedia retrieval, and argue that techniques such as thesauri will be needed for the foreseeable future in order to support users in their need for multimedia content. In particular we argue that Artificial Intelligence (A.I.) techniques are not mature enough to solve the problem of indexing multimedia conceptually, and will not be able to replace human indexers for the foreseeable future.

1. INTRODUCTION

This special issue focuses on the argument between those who posit that the traditional thesaurus has no place in modern information retrieval, and those who say it does. In this position paper we argue the latter - thesauri do have a place in modern information retrieval for multimedia content, and will do for the foreseeable future. Knowledge organisation in general plays a significant role in multimedia search, providing human indexers with meta-data schemes (i.e. schemes that list elements of a multimedia document such as concept, theme, people etc) together with ontologies or thesauri with which the multimedia document can be tagged or labelled. In this paper we set out the reason for this – namely the semantic gap problem in multimedia (section 2), point to some current schemes used to indexing multimedia content (section 3), and argue that alternatives to manual indexing are a long way off due to little understood and hard to tackle computational problems (section 4). We provide a conclusion in section 5.

2. THE SEMANTIC GAP IN MULTIMEDIA INFORMATION RETRIEVAL

Text information retrieval has become very successful by automatically indexing content of documents by the words contained within them. It is a straightforward issue to identify semantically meaningful content via keywords, and to use statistical techniques to rank text documents according to their relevance to the user (Robertson et al., 1995). User requests in text retrieval are easy to implement in software and have been shown to work for the user.

Multimedia (documents which contain either images, sound or streams or both and may also contain text) unfortunately cannot be treated this way. Video and speech audio can produce transcripts via technologies such as speech recognition techniques (e.g the Informedia project (2009) at Carnegie Mellon University), which although full of errors can be used for indexing multimedia documents. However without any terms to associate with a multimedia document as with images or music, there is an inherent problem indexing such objects. It is possible that the document

has some meta-data associated with it, but this is not always the case (e.g. on the web). With the web, multimedia documents are becoming increasingly more readily available, and mechanisms to access such information are sorely required.

A proposed solution to the problem is to use Content Based Information Retrieval (CBIR) methods, and index the multimedia document by its underlying low level content (Rueger 2009). For images this would be colour, shapes, texture etc., (Rueger 2009, 44) and for music key, tempo, harmony etc. (Chowdhury 2004, 302). For some applications, this technology works well e.g. in pattern/design matching, artwork textural analysis, trademark matching and music services such as Shazam. However, the key problem in many applications is that these low level features do not match high level concepts, and therefore CBIR technologies have had only limited success. This problem is known as the *semantic gap* – this can be defined more formally as - paraphrasing Enser (2008a, 537) - the "rift in the information retrieval landscape between the information that can be extracted automatically from a digitized object and the interpretation that humans might place upon the object". We argue in this paper that content-based technological solutions to multimedia retrieval are a long way off due to this semantic gap, and that knowledge organisation techniques such as thesauri will be required for many years to come (an argument developed further in section 4).

3. USE OF KNOWLEDGE ORGANISATION IN MULTIMEDIA INFORMATION RETRIEVAL

In order to support retrieval of Multimedia documents, professional indexers with subject knowledge are required, who have access to appropriate knowledge organisation tools such as metadata schemes, domain ontologies and thesauri. These manually address the semantic gap problem described in section 2. It should be noted that thesauri are not the total answer to the challenges in the area, and should be understood in the context of other techniques such as uncontrolled indexing and non-KO methods. We point to some of the many freely available knowledge organisation schemes for images and music here – for a more comprehensive view see the Riley and Becker (2010) visualisation of meta-data standards.

There are a number of specialist tools which can be used for indexing images, including the Library of Congress's *Thesauri for Graphic Materials* (TGM) and *Iconclass* maintained by the Royal Netherland Academic of Arts and Sciences. *TGM* (Library of Congress N.D.b) is a more generic scheme and covers all kinds of graphical media including photographs, prints, paintings and drawings etc. Indexers can choose terms based on objects in the image, relationships between those objects, choose broader or narrower terms, establish syntax, and refer to notes which give context of use etc. *Iconclass* focuses specifically on Art images, providing the indexer which three general areas to choose from – abstract art, general division (religion and magic, nature, humanity) and specific divisions (history, bible, literature). Each scheme allows the indexer to assign search terms to images for the purposes of retrieval. Other schemes include the Visual Resources Association Core Categories (VRA N.D) and the Categories for the Description of Works of Art (Getty 2014). The Getty research institute has a number of very useful tools including the Art & Architecture Thesaurus (Getty N.D.a) and the Thesaurus of Geographic Names (Getty N.D.a) which are line (these are prominent in the linked data community, an issue dealt with in another paper in this special issue). Somewhat surprisingly there are few thesauri for moving images e.g. videos, apart from that provided by the Library of Congress (Taves et al., 1998), forms/genres for films/videos (Library of Congress N.D.a), the FIAF cataloguing rules for film archives

(Harrison 1991) and the Multimedia Content Description Standard MPEG-7 (MPEG-7 N.D.). A fuller review of schemes can be found in (Enser 2008b).

The Library of Congress has also been very active in the music domain and has for a number of years been in the process of developing music/genre headings (Library of Congress 2013a; 2013b) and performance terms for printed music (Library of Congress 2013c; 2014d), largely (but not exclusively) for the domain of Western Classical music. This development is welcome given the general dissatisfaction with the lack of standard music thesauri in the library community, and lack of work in recent years by the musicology community on the problem. Apart from this there is very little work in the area, apart from the British Catalogue of Music Classification (Coates 1960), which was abandoned by the British Library in 1982, being replaced for the most part by Dewey Decimal Classification (DCC N.D.).

It can be seen from the above selection of schemes that knowledge organization techniques are still under very active development and use in sectors that make use of images, audio materials and multimedia, but what of future developments? Let us consider this matter next.

4. THE FUTURE OF KNOWLEDGE ORGANISATION IN MULTIMEDIA INFORMATION RETRIEVAL

So what of the prospects for the use of computers to automatically index multimedia content? What is required in terms of human subject knowledge by software to be able to carry out such an activity and either replace or augment humans?

A proposed solution to the problem is to use Artificial intelligence (A.I.) techniques, as proposed by Alan Turing (Turing 1950). A.I. can be defined as the use of computers to solve given problems, such that it would be impossible to tell the difference between a human and a program as to who carried out the work for the solution – proposed in a thought experiment by Turing (Turing 1950) called 'The Imitation Game'. An example would be to take an image, get a human and a computer programme to index a multimedia object, and see if a third party human could tell which solutions were provided by the human and which by the computer. If we could get such technology to work, we could deploy a software programme to index images, music etc without the need to use human indexers, and perhaps we could reach the stage where we could offer the same service for multimedia retrieval as we can for text retrieval. This has been the goal of many working in multimedia search for a number of years. There has been some success in terms of identifying objects in images (Karpathy and Li, 2015), but conceptual issues are much harder to address (see below).

Clearly this has not been totally successful to date, so what are the barriers to success in the use of A.I. technologies to solve multimedia indexing problems? To understand this we need to understand the knowledge that an indexer has built up over years in the subject in which they are working. For example a human indexer needs to understand the concept of 'Genre' in music and images, the cultural context in which they are used both in terms of space and time (e.g. a renaissance painting with an aristocratic Italian woman as the subject). The indexer needs to build up a significant body of tacit knowledge (Polanyi 1966), both in terms of the subject itself and the process of indexing (e.g. many years of engaging with renaissance literature and art to interpret Italian renaissance painting). For example, in image retrieval the indexer needs to understand both the 'of-ness' and 'aboutness' of an image (e.g. what is the context for the painting of the Italian woman – who was she and why was the portrait commissioned?). This knowledge is experiential and is hard to pass on to other humans, let alone software products.

A.I. has promised much over the years, but has delivered results only slowly and incrementally. Far too many people have expected A.I. to deliver rapid and significant results quickly, and the hype surrounding it has often led to disappointment, which in turn has led to 'A.I. winters' where funding for research has been cut significantly. This has been to the detriment of the field and to progress in solving complex computational problems, and therefore to the application we address here – multimedia IR.

We turn to the use of A.I. techniques specifically for image retrieval as a technology which can be applied to multimedia IR. Enser (2008a) relates the failure of CBIR systems based on A.I. techniques to fulfil their promise, with critics in the information science community demonstrating through experimentation that users do not find low level features useful for search. Commercial systems that proposed using such an approach have failed to make any headway as a result. A focus on the more semantic aspects of the content (Enser, 2008a) has proved to be problematic also as some concepts in an image are intrinsic to it and are not physically present (e.g. a picture of a politician involved in an election – the 'politician' is identifiable, but the concept of an 'election' is more difficult to detect). Technologies which detect these intrinsic concepts or 'aboutness' do not currently exist – this is the core problem in the field (see examples above). The use of ontologies has been proposed, but this requires significant user input to build the ontology and does not get us nearer to the process of automating indexing an image without human intervention. As (Enser 2008a, 539) argues "At the present time, most attempts at bridging the semantic gap have faltered at the very broad separation between object labeling and the high-level reasoning which situates those objects appropriately within the viewer's sociocognitive space". The challenges in the area are significant and solutions will be a long time coming – the semantic gap is here to stay (for a while at least!).

5. CONCLUSION

A.I. technology has come a long way over the past 60/70 years when it was first proposed, but it has not yet built up the capacity to deal with problems in knowledge for indexing multimedia documents. This still requires significant human input as argued in section 4, and it is unlikely that we will see computing solutions to the semantic gap in the area any time soon. It is impossible to see the future, and we cannot rule out the possibility that some technology will eventually come along and replace the indexer – although as Carr (2015) has pointed out, simply automating a problem does not automatically solve all the issues which arise, and may in fact introduce more (e.g. confirmation bias, automation bias etc). Some progress has been in detecting specific objects in images, to generate image descriptions (Karpathy and Li, 2015), but these still do not address the conceptual problem outlined in section 4. There may be some mileage in a hybrid approach e.g. building appropriate ontologies, which can then be used by machine learning algorithms to categorise and classify images, using the features extracted by CBIR algorithms (e.g. a semi-supervised learning approach).

It is our view that practitioners in knowledge organisation generally and thesauri in particular will be needed to support multimedia information retrieval for the foreseeable future i.e. human indexers to classify or tag multimedia objects so that the user can find them in search.

REFERENCES

Carr, Nicholas. 2015. *The Glass Cage: how computers are changing us.* London: The Bodley Head.

Chowdhury, Gobinda G. 2004. *Introduction to Modern Information retrieval: 2nd Edition.* London: Facet Publishing.

Coates, Eric J. 1960. *The British catalogue of music classification.* London: Council of the British National Bibliography.

Dewey Decimal Classification. N.D. Introduction to Dewey Decimal Classification. http://www.oclc.org/content/dam/oclc/dewey/versions/print/intro.pdf .

Enser, Peter G.B. 2008a. The evolution of visual information retrieval. *Journal of Information Science,* 34: 531-546.

Enser, Peter G.B. 2008b. *Visual image retrieval.* In *Annual Review of Information Science and Technology* , edited by Blaise Cronin. Chichester: Wiley. 42: 3-42.

Getty. 2014. Categories for the Description of Works of Art website. http://www.getty.edu/research/publications/electronic_publications/cdwa/.

Getty. N.D.a. Getty Art & Architecture Thesaurus® online website. http://www.getty.edu/research/tools/vocabularies/aat/.

Getty. N.D.b. Getty Thesaurus of Geographic Names® online website. http://www.getty.edu/research/tools/vocabularies/tgn/.

Harrison, Harriet W. 1991. *The FIAF Cataloguing Rules for Film Archives.* München: K. G. Saur.

Iconclass. 2012. Iconclass website. http://www.iconclass.nl/.

Informedia Project. 2009. Informedia project website, Carnegie Mellon University. http://www.informedia.cs.cmu.edu/.

Karpathy, Andrej and Fei-Fei Li. 2015. Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of CVPR 2015, Boston, 7-12 June 2015,* edited by Kristen Grauman, Eric Learned-Miller, Antonio Torralba, and Andrew Zisserman. http://cs.stanford.edu/people/karpathy/deepimagesent/.

Library of Congress. 2013a. Genre/Form Terms for Musical Works and Medium of Performance Thesaurus. https://www.loc.gov/catdir/cpso/genremusic.html.

Library of Congress. 2013b. Genre/form terms agreed on by the Library of Congress and the Music Library Association as in scope for Library of Congress Genre/form terms for Library and archival materials (LCGFT). http://www.loc.gov/catdir/cpso/lcmlalist.pdf.

Library of Congress. 2013c. Introduction to Library of Congress Medium of Performance Thesaurus for Music. http://www.loc.gov/aba/publications/FreeLCSH/mptintro.pdf .

Library of Congress. 2013d. Performance terms: Medium. http://www.loc.gov/aba/publications/FreeLCSH/MEDIUM.pdf.

Library of Congress. N.D.a. Library of congress genre/forms for films video. http://www.loc.gov/aba/publications/FreeLCSH/GENRE.pdf.

Library of Congress. N.D.b. Thesaurus for Graphical Materials (TGM). http://www.loc.gov/pictures/collection/tgm/.

MPEG-7. N.D.. Moving Pictures Expert Group MPEG-7 standard website. http://mpeg.chiariglione.org/standards/mpeg-7.

Polanyi, Michael. 1966. The Tacit Dimension, Chicago, Il.: University of Chicago Press.

Riley, Jenn and Dawn Becker. 2010. Seeing standards: A Visualization of the Metadata Universe. http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards.pdf .

Robertson, Stephen E., Stephen Walker, Susan Jones, Micheline Hancock-Beaulieu and Michael Gatford. 1995. Okapi at TREC-3. In *Proceedings of the third text*

retrieval conference, Gaithersburg, November 1994, edited by Donna Harman, NIST SP 500-226. Gaithersburg, Dc.:NIST. 500-226. 109–126.

Rueger, Stefan. 2009. Multimedia Resource Discovery. In *Information Retrieval: Searching in the 21st Century*, edited by Ayse Goker and John Davies. Chichester: Wiley, 39-62.

Taves, Brian, Judi Hoffman and Karen Lund. (1998). The Moving Image Genre-form Guide. http://www.loc.gov/rr/mopic/migintro.html.

Turing, Alan. 1950. Computing Machinery and Intelligence, Mind 49: 433-460.

VRA. N.D. Visual Resources Association Core Categories website. http://core.vraweb.org/.