



CERLIM Centre for Research in Library
and Information Management

**Synchronised Object Retrieval:
a feasibility study into enhanced
information retrieval in multimedia
environments using synchronisation
protocols**

Library and Information Commission Research Report 92

Peter Brophy
Richard Eskins
Tony Oulton

December 2000

**CENTRE FOR RESEARCH IN LIBRARY & INFORMATION MANAGEMENT
DEPARTMENT OF INFORMATION & COMMUNICATIONS
THE MANCHESTER METROPOLITAN UNIVERSITY**

Synchronised Object Retrieval: a feasibility study into enhanced information retrieval
in multimedia environments using synchronisation protocols
(Library and Information Commission Research Report 92)
by Peter Brophy, Richard Eskins and Tony Oulton
The Manchester Metropolitan University, 2000.

© Copyright Resource: The Council for Museums, Archives and Libraries 2000.

The opinions expressed in this Report are those of the authors and not necessarily
those of Resource: The Council for Museums, Archives and Libraries.

RE/084

ISSN 1466-2949

ISBN 1 9023 9453 4

The authors have asserted their Moral Rights.

CONTENTS

Executive Summary.....	3
1. Introduction	4
2. The Synchronized Multimedia Integration Language (SMIL)	5
3. Review of Related Work.....	7
3.1. Text Retrieval.....	7
3.2. Audio Retrieval.....	9
3.2.1. Sound retrieval (non-verbal audio).	10
3.2.2. Speech retrieval	11
3.3. Image Retrieval.....	12
3.3.1. Content-based Image Retrieval – definitions etc.	13
3.3.2. Research Context	13
3.3.3. Users – needs, searches, satisfaction.....	13
3.3.4. Indexing.....	15
3.3.5. Non-verbal interfaces.....	18
3.3.6. Search process	18
3.3.7. Search engines for image retrieval	19
3.3.8. Information retrieval from videos	20
3.4. Metadata, standards	22
3.5. Combined approaches and the use of synchronisation	23
4. Experimental Design.....	24
4.1. Establishment of test collection.....	24
4.2. User input	25
4.3. Query Analysis	25
4.4. Search of packages	27
4.5. Search of micro-objects' metadata	27
4.6. Search of micro-objects' content	27
5. Conclusions	28
6. References.....	29
Appendix: Synchronized Multimedia Integration Language (SMIL) A WWW	
Bibliography	38
A. HyperText Markup Language @ W3C	38
B. SMIL @ W3C	38
C. Microsoft.....	39
D. Demos/Tutorials	39
E. Other Multimedia Projects.....	41
F. SMIL Web Sites and Lists of SMIL Resources	41
G. SMIL Mail Lists	42
H. SMIL Articles	42
I. SMIL Software.....	45
J. Examples.....	46

Acknowledgements

The authors would like to acknowledge the critical input of colleagues to this work, and in particular the assistance of Dick Hartley and Frances Johnson in the analysis of the state of the art of information retrieval research.

The financial support of Resource: the Council for Museums, Archives and Libraries (and initially the Library & Information Commission) is gratefully acknowledged.

The views expressed in this Report are those of the authors, who are responsible for any errors or omissions.

PB
RE
AJO

December 2000

Executive Summary

The research reported here investigated the feasibility of enhancing the retrieval of both multimedia presentations and mono-media 'micro-objects' within such presentations by utilising information contained in synchronisation files. Using the widely-accepted *Synchronized Multimedia Integration Language* (SMIL) standard, the work examined the features of compliant multi-media packages which make enhanced retrieval possible. The particular and unique feature of this approach would be that the retrieval 'score' of any micro-object would be modified by considering the scores of all other micro-objects designed to be played in parallel with it. Where both gave positive results, an inference could be drawn that the object being examined was more likely to be relevant than could be concluded from examining the object individually – and of course the converse would be true.

The work included an extensive review of the literature of mono-media information retrieval (including text, audio and image retrieval) and examined the relatively small literature which refers to multimedia approaches. The conclusion was that the approach being examined is indeed unique.

A outline experimental design was developed which noted that the further development of the approach would need major research effort, not only to prove the concept itself but to develop a suitable SMIL-compliant collection which could be used to test different experiments, different algorithms and different weightings of results.

Overall, the research concluded that the *Synchronised Object Retrieval* approach is both feasible and of potentially enormous significance in networked, multimedia information environments.

1. Introduction

This Report contains the results of a feasibility study into the use of synchronisation protocols and data, and in particular the SMIL (*Synchronized Multimedia Integration Language*) standard (World Wide Web Consortium, 2000), as a basis for enhanced retrieval of mono-media objects within multimedia domains – we have termed this *Synchronised Object Retrieval (SOR)*. The idea behind this research is that because SMIL defines a temporal relationship between objects within a multimedia presentation, it should be possible to extract retrieval clues from the simultaneity of playback of constituent ‘micro-objects’ (this term is used to describe any mono-media file which is a component of a multimedia presentation). For example, if it is known that a text file is to be played back at the same time as an image file, and a text-based search of the text file retrieves relevant keywords, then if the same query expressed as a content-based search of the image file also produces possible matches, the inference can be drawn that

- the multimedia object as a whole has relevance to the query
- the instance identified (i.e. the simultaneous display of the text and image files) has relevance
- the text file has relevance
- the image file has relevance

and each of these conclusions may be drawn with a higher degree of confidence than a simple search of any one component part of the multimedia presentation could produce. In other words, the known synchrony of the files allows reinforcement of retrieval conclusions.

Since it is likely that an increasing amount of information, within an ever expanding information universe, will be delivered in multimedia formats in the future, such techniques could be of great significance. They could be used either

- to enable the retrieval of individual files from within multimedia presentations across collections of heterogeneous multimedia objects
- to enable the retrieval of multimedia presentations from one or more collection of such presentations, by examining the results of searching across some or all of a presentation’s constituent ‘micro-objects’.

There are many possible applications of such capability – an example would be that of a lecturer searching across a large collection of multimedia objects to find micro-objects to re-use in a new presentation.

It has to be admitted that searching in this way will require a very high level of computing power. While in the past this may have been a reasonable objection to the SOR approach, the effects of Moore’s Law (the inexorable doubling of computing power every eighteen months) suggests that computing power may not be an issue in the future. Indeed, a recent paper by Arms (2000) makes the observation that “simple algorithms plus immense computing power often outperform human intelligence” and points out that processes which in the past would have been unthinkable are now routine, using the example of the Google Internet search engine:

“Evaluating the importance of documents would appear to be a task that requires human understanding, but Google’s ranking algorithm does remarkably well entirely automatically Calculating the ranks requires the algorithm to iterate through a matrix that has as many rows and columns as

there are pages on the web, yet with modern computing and considerable ingenuity, Google performs this calculation routinely.”

This Report documents the research undertaken in a short feasibility study, funded by *Resource: The Council for Museums, Archives and Libraries*, which was designed to investigate the SMIL standard itself and the state-of-the-art in the IR fields which might contribute to SOR, and then to outline an experimental design which might be implemented to test the SOR approach in an experimental IR environment. It is the intention of the SOR research team to pursue funding to enable this next phase of research to be undertaken. However, the scope of this work takes it beyond the current remit of Resource and hence funding is being sought from other sources. In the meantime the research team would welcome input and observations from any interested parties.

2. The Synchronized Multimedia Integration Language (SMIL)

The *Synchronized Multimedia Integration Language* or *SMIL* (pronounced ‘smile’) standard is based on XML (the eXtensible Markup Language) and its syntax is defined in an XML DTD (document type definition). SMIL documents can be authored using a simple text editor, since they are in essence similar to HTML, although a variety of SMIL-enabled authoring tools are available. The project team investigated these and used the *Real Slideshow Plus* package as a basis for exploring the potential of SMIL (a demonstration SMIL application is available at <http://www.mmu.ac.uk/h-ss/cerlim/projects/synchro/demo.htm>). At the basic level such applications enable an author to create a series of ‘micro-objects’ as part of a presentation or ‘macro-object’ and to define the synchronisation over time that is required. The standard defines a number of different types of media micro-objects that can be included in a presentation.

animation: animated vector graphics or other animated format

audio: audio clip

img: still image

text: text reference

textstream: streaming text

video: video clip

ref: generic media reference (in effect, ‘other’)

For example, an author can write synchronisation instructions such as: ‘play audio file A in parallel with video file B’ and ‘display text file C after audio file A and in parallel with animation file D and with text stream E’. In addition SMIL enables the author to define the positioning of micro-objects on the user’s screen (or ‘visual rendering surface’ in the terminology of the standard). Various other variables can be used, such as the ability to test the end-user’s bandwidth and deliver alternative micro-objects accordingly. However, these features are unlikely to be of major significance in the context of an information retrieval system.

As part of the feasibility study the team undertook a detailed evaluation of mainly web-based resources relevant to SMIL. The results of this work are available (with

clickable links) on the project web site at <http://www.mmu.ac.uk/h-ss/cerlim/projects/synchro/smil-bib.htm> and are reproduced as Appendix 1 to this report. We have provided commentary on each of the sources selected. Readers who are unfamiliar with SMIL and the construction of SMIL-compliant presentations may like to work through one of the demonstrations available, such as that at <http://www.helio.org/products/smil/tutorial>

One of the advantages of the SMIL approach is that it effectively enables objects which would normally require high-bandwidth, such as multimedia packages, to be delivered across low-bandwidth networks since each constituent micro-object (file) can be transmitted separately and the package correctly reassembled and played at the client end, using the synchronisation information. It thus has enormous potential for the distribution of all kinds of multimedia objects. Furthermore, there is no reason why the micro-objects cannot be repackaged and delivered within separate presentations, without the expense of re-recording a video or other complex element as might be the case were the presentation to be recorded as a single file of data. Among other useful aspects of the approach is the use of alternative files: for example, a commentary can be stored in a variety of languages and the user given the option as to which should be played (or the option picked up automatically from the user's preference settings). SMIL 2.0 (the current standard) has introduced the concepts of *modularisation* and *profiling* to enable functionality to be extended and the integration of related mark-up languages (an example might be MathML)

The standard enables the author to define not only the temporal order in which the micro-objects should be played but also the spatial layout of the presentation. For example, generic presentations can be defined in terms of layout on a standard computer screen, although it is also possible to define device-dependent layouts.

There are a number of interesting features of SMIL from an information retrieval standpoint, of which the following are examples.

- ◆ The standard defines an RDF compliant 'metainformation' module which should enable the macro-object and its constituent micro-objects to be retrieved like any other documents by examining metadata elements. While the schema used is open, the SMIL 2.0 standard suggest the use of Dublin Core as a generalised approach and an example of this usage can be found in section 8.5 of the standard. In addition to, say, subject-based retrieval (using the DC *subject*, *title* and *description* elements for instance) a SOR system could use some elements to refine search results (using the DC *creator* element to find micro-objects which were created by the same person so as to infer some commonality might be an example).
- ◆ SMIL supports the inclusion of linked objects in a presentation, through its *Linking Module*, enabling events (such as a user clicking on an area of the screen) to trigger a link to another part of the presentation. As in HTML, provision is made for supporting name fragment identifiers (using the # symbol) so as to enable a link to be made to any tagged element within a micro-object. (For example, if the SMIL 2.0 documentation was itself part of a SMIL presentation, a link could be provided to <http://www.w3.org/TR/smil20/extended-media-object.html#edef-ref> where definitions of media object elements would be found.) If this is not the start of the micro-object the effect would be as if the user had fast-forwarded within that object to a specified place. Further refinement is possible: for example, SMIL contains provision to enable the author to specify precedence levels for cases where the origin and linked objects continue to

play simultaneously (an example would be setting the relative audio playback levels). Again, although we have not explored these in detail in the feasibility study, it would be possible for software to follow embedded links and to take account of associated variables to retrieve further micro-objects which may themselves contain both usable metadata and other retrieval clues.

- ◆ The SMIL 2.0 standard has specified various features intended to improve accessibility. For example a 'long description' attribute enables the author to provide a long text description (separate from metadata), which could be very useful for enhancing retrieval. Because of the ability to define user-required variants it is also possible to allow for the identification of, for example, text equivalents of audio intended for closed-captioning, thus providing an accessibility feature and further IR possibilities.
- ◆ As noted above a 'language' attribute is available, intended for presentations where the user has choice of, say, a French, German or English sound track – the multimedia package would of course contain three files, each synchronised with the video and other elements. Again, the possibilities for enriching IR capability (by combining searches in a variety of languages) are intriguing.

3. Review of Related Work

This review of the literature is intended to outline significant developments, particularly research activities, related to information retrieval from single media (text, sound and still images) and multi-media sources (primarily moving images in the form of videos) from the mid-90s. Sources reviewed are all in English with a predominance of material from North America and Europe. As some authors indicate, this corpus does not always sufficiently reflect significant work in other languages, notably Japanese. The review is intended to highlight developments in information retrieval from the various sources, which might be used to support retrieval of individual components of 'presentations' through the use of synchronisation protocols. It is therefore indicative of current interests and concerns, rather than exhaustive.

Text retrieval is reviewed first, as the fundamental to many other developments, followed by audio retrieval. Audio retrieval problems are often treated in a manner similar to text, because of the significance of verbal forms. Retrieval of images presents other problems particularly information retrieval from video, which is a major strand in research and development. Information retrieval from video incorporates ideas and techniques from the other forms of retrieval – image, audio and text. Work on combined approaches, including synchronisation, is described in the final section.

3.1. Text Retrieval

The most significant development in text retrieval in the recent past has not been new understanding or technical achievements. Rather it has been the application of earlier research and development in text retrieval to large databases, often unstructured, with a common intention being to facilitate access not just by end-users but by what Travis (1998) called 'casual users' – relatively naïve and infrequent users. In such circumstances, retrieval systems using statistical and probabilistic search engines or natural language processing for both query and search are better suited to relatively unskilled searching, particularly when databases, because of

sheer size or economic constraints, are not well organised. In these circumstances, Boolean search capabilities, which have worked well when used by trained searchers with well-indexed documents (Feldman, 1999), are less useful, although Marchionini, Dwiggins et al. (1993) have suggested that there is a trade-off between domain knowledge and search skills. Boolean logic is explicitly available or used implicitly in many Web search engines.

Enhanced user input to the indexing process may be a cost-effective method of improving indexes to large heterogeneous databases. Desai, Singhal et al. (1999), describe experiments with an expert system (CINDI – Concordia INdexing and Discovery System) to enhance the quality of user input in the indexing process. CINDI is intended to support the creation of metadata in the form of semantic headers at a level of professional cataloguing by the originators of Web documents, and to facilitate searching by naïve users as though by a reference librarian. A key aim is to compensate for Web search engine indexing limitations. The CINDI input interface provides assistance with the application of a controlled vocabulary and allows subsequent users to add annotations, which could be used, for example, for peer review comments.

The primary example of a very large, unstructured information resource is the World Wide Web. The title of Travis's editorial is *From "storage and retrieval systems" to "search engines"* – a concept echoed by Feldman (1999) in her review of the current state of text retrieval. An overview of web search engines, including metasearchers, intelligent agents etc., is provided by Chung and Lee (1998). In a preliminary report of a study by Bourne and Hahn of the early development of online systems Hahn (1998) shows that all 24 key developments in web searching were the outcome of research and experiments in the mid-60s to early 70s. Key concepts of system performance measurement, notably recall and precision, were also developed in this period. Other performance criteria such as relevance and usability – defined by Feldman (1999) as speed, interface design, opportunities to interact with the system and the ability to use natural language queries – will become increasingly important in the context of open access to electronic searching. Tools to support relevance judgements, such as visualisation, and question-answering rather than document identification and retrieval are likely to be significant in this process. Chiaramella (1997) contrasted the notions of *standard* system relevance with *particular* user relevance within the context of interactive retrieval performance, stressing the importance of *real* integration of traditional IR querying and hypermedia browsing facilities. Greene, Marchionini et al. (2000) discuss the use of 'previews' and 'overviews' in the design of interfaces to digital resources as a means to facilitate easy identification of relevant items.

Web search engines now provide, although not necessarily in one engine, all the search capabilities for information retrieval – explicit or implicit use of Boolean operators, exact phrase searching, truncation (and related search techniques, such as stemming and masking), thesauri, full text searching and relevance ranking of hits. The commercial development of web search engines has negative consequence in the secrecy surrounding relevance ranking algorithms, for example, and ongoing developments.

A significant development at the input end has been the ease with which text in electronic format can be created and made available for searching, from word-processed rather than typed documents, and the relative simplicity of creating and mounting web pages. One recent estimate from Inktomi puts the number of unique documents on the web at 500 million, and a similar number of duplicates etc. (Brewer, 2000). It is the stated intention of Inktomi to index **all** (our emphasis)

meaningful content, including newer formats such as MP3 and video. In such an environment Oppenheim, Morris et al. (2000) argue for the development of a standardised set of tools for web search engine evaluation.

A significant development in text retrieval is claimed to be new theoretical bases for text retrieval, for example the work of embodied in the Readware technology (Ewell and Adi, 1997). The Dempster-Shafer theory may offer an alternative to standard probabilistic models (Jose and Harper, 1997), and has been applied to qualitative data (Parsons, 1994). Ghosh and Chaudhury (1998) presented an argument for using abductive logic to determine the relevance of documents for a query. Van Rijsbergen (1999) is currently exploring the use of this model of reasoning to provide a framework for an explanation-based account of relevance feedback. Arms (2000), on the other hand – as we have seen - supports the observation that ‘simple algorithms plus immense computing power often outperform human intelligence’. His argument, that an increase in computing power by two orders of magnitude in the next decade (based on historical trends) will reduce the need for information skills, may perhaps be extended to intellectual developments.

Data fusion - retrieving information from several sources at once and merging the results into a single response set - can be seen to be a major trend in research and applications of information retrieval. Autonomy claims that its pattern matching software combined with the SoftSound speech recognition system allows completely automatic management and searching of both text and speech (Autonomy, 2000).

Such developments will begin to address or redefine the current areas of research and development in text retrieval identified by Feldman (1999): file structures (large inverted or smaller signature file structures, or relational databases); speed of operation; limited field or whole text searching; usability; the problems of polysemy and synonymy; indexing consistency; spelling consistency and typographical errors, including the limitations of OCR systems; precision and false drops.

Since 1991 the series of TREC (Text REtrieval Conference) workshops has been organised to support text retrieval research in a ‘real life’ context, by providing large test collections, uniform scoring procedures and a forum for researchers to discuss and compare results. The main, fundamental, task is testing the performance of retrieval systems on a static set of documents using changing questions. The tracks (that is supplementary tasks) have focussed on related research, such as retrieving documents written in a variety of languages using questions in a single language (cross-language retrieval); retrieving documents from very large (100GB) document collections; and retrieval performance with humans in the loop (interactive retrieval) (Voorhees and Garofolo, 2000).

3.2. Audio Retrieval

Audio retrieval, especially music and speech, is an active research area although still in its infancy (Goodrum and Rasmussen, 2000). Foote argued in 1998 that it was then entirely possible to locate audio of someone *talking* about Martin Luther King, or indeed one of his speeches. Information retrieval from audio sources is commonly discussed in terms of verbal (speech retrieval) and non-verbal elements (sound retrieval), although there are obvious overlaps.

Before considering these areas it is worth noting the use of sound (*sonification*) as display device, akin to and sometimes coupled with visualisation. Such sound

patterns can be equally well-detected and described by musically and non-musically trained people (Gluck, 2000).

3.2.1. *Sound retrieval (non-verbal audio).*

One requirement for a sound retrieval system might be to retrieve sounds which match a given sound in some characteristic(s) – the sound of applause, or a sound ‘like the sound of a herd of elephants’ (Wold, Blum et al., 1999). Retrieval of sounds in this way would be useful by, for example, sound effects engineers. Foote (1998) cites an example of the detection of noises such as shots, cries and explosions to identify violence in films automatically from the soundtrack. Retrieving sounds has much in common, e.g. fuzzy matching via similarity measures, with image retrieval and pattern recognition (Wold, Blum et al., 1999). Wold, Blum et al. also argue that indexing of acoustic features is necessary to facilitate retrieval from a large database. Yoshitaka and Ichikawa (1999) suggest that content-based retrieval of audio data is less developed than still and moving image studies, the key reason being the difficulty of extracting unique identifiers from individual sound sources.

Different researchers, e.g. Berlin Technische Universität, Muscle Fish and Foote, use varying approaches to audio retrieval-by-content. Their different approaches tend to favour one characteristic of sounds, for example pitch or timbre. In part this is due to the subjective nature of defining ‘similarity’ in sound retrieval (Foote, 1998).

Several researchers in the retrieval of music use text retrieval systems in a challenging manner, for retrieving *musical words* (n-gram note sequences) contained within melodies (Goodrum and Rasmussen, 2000). Ghias, Logan et al. (1995) demonstrated one successful use of this approach, based on the relative differences of pitch of three successive notes (a form of trigram based on Up, Down, Same) automatically generated from a sequence of hummed notes. They used these trigrams to retrieve songs similar to the hummed melody from a database of 183 songs in MIDI format. The method seemed robust enough to cope with varied musical quality in the humming. They suggest that higher resolution (more notes and /or more pitch discrimination between adjacent notes) would be necessary for larger databases. They also suggested that the use of wavelets would improve the efficiency of the pitch-tracking algorithm. Wavelet transformations of sound differ from Fourier analysis in providing time and frequency localization properties (Li, 1999?). Downie also used the idea of n-gram musical words in his research, but for interval-only representations of monophonic melodies (Downie, 2000b). Using traditional text retrieval techniques, Downie reports precision scores of 97-99% in some of his tests. The work of David Bainbridge and colleagues has focussed on the development of a system which combines melody-based querying with optical recognition for music scores, as well as text searching, for use with the 100,000 MIDI files within the New Zealand Digital Library of Popular Music (Bainbridge, Nevill-Manning et al., 1999; Missingham, 1999).

Blackburn and DeRoure (1998) had as their focus content based navigation of music by hypermedia linking from one file to another, and between files. Their orientation is within Southampton’s Microcosm and MAVIS programme which is discussed later. Working with a database of 8,000 songs they conclude that, for their purposes at least, the contour representation approach of query-by-humming is not sufficiently accurate and suggest modified and alternative approaches.

The OMRAS (Online Music Recognition and Searching) Project is a 3-year collaborative effort between King’s College London and the Center for Intelligent Information Retrieval at the University of Massachusetts funded by JISC and the National Science Foundation under the Digital Libraries programmes, from 1999.

The aim is to explore methods for retrieval of music in a variety of forms - encoded score-data, MIDI and similar files and digital audio recordings (OMRAS, 1999). JISC also supported the MuTaTeD! (Music Tagging Type Definition) Project at Glasgow University aimed at developing a meta-standard for music mark-up (Boehm, 2000; <http://www.pads.ahds.ac.uk/mutated>) which has been continued as MuTaTeD'II with LIC funding, aimed at designing and implementing a music information retrieval system for encoded music (<http://pads.ahds.ac.uk/MuTaTeD2home.html>).

The first International Symposium on Music Information Retrieval is due to be held in October 2000, with a varied programme including music representation and indexing, evaluation of music-IR systems and user interfaces for music IR (Downie, 2000a). A significant factor in the establishment of this Symposium was the *Exploratory Workshop on Music Information Retrieval* organised at the ACM SIGIR '99 conference – abstracts are available at URL: http://lis.uiuc.edu/~jdownie/mir_papers/sigir99_wshop_proc.pdf.

3.2.2. *Speech retrieval*

Speech retrieval has been the subject of research from two directions - speech recognition (essentially transcribing spoken words to text), and sound recognition (identifying the meaning of spoken documents without transcription). Automatic Speech Recognition (ASR) is the more developed and is a significant element in a number of projects, such as Informedia at Carnegie-Mellon.

Existing technology in reasonable circumstances is able to determine speaker identification very accurately (Foote, 1998). Identifying changes between speakers can provide a useful device for segmenting audio recordings, and associated video or other data streams. One such operational application was the matching of the corrected and emended text of the US *Congressional Record* with original voice recordings to see what a speaker actually said (Foote, 1998).

Foote (1998) argues that a 'perfect' ASR system would essentially reduce speech retrieval to known text retrieval problems. He suggests that while ASR systems are far from perfect, with accuracy ranging from 90% in particular circumstances to 50%, they are still surprisingly helpful in information retrieval. Singhal and Pereira (1999) argue that the use of document expansion (their term for enhanced document representation) in automatic speech recognition results in retrieval effectiveness which is indistinguishable from retrieval from person-made transcripts if the automatic transcription error rate is less than 25% and the document expansion is based on a relevant corpus. Witbrock and Hauptmann (1997) evaluated the effect of different transcription error rates on retrieval, by interpolation from perfect transcripts, and found that at word error rates less than about 25% the performance of the best search engine (an enhanced version of the basic search engine) is very close to perfect text transcriptions.

Spoken document retrieval (SDR) has been a track (research programme) within the Text Retrieval Conference since TREC-6, 1997. The research consists of producing time-marked transcripts by speech recognizers of audio broadcast news stories. The transcripts are then indexed and searched by information retrieval systems with performance criteria set by searches of human transcriptions (the 'ground truth' transcripts). Queries (called 'topics' in TREC terminology) are determined by NIST (the National Institute of Standards and Technology). The performance of the best automatic systems was just less than 90% of searches of human transcripts. The SDR test database has grown from 1,451 stories in 50 hours of broadcast in 1997 (TREC-6) to almost 22,000 stories in 557 hours in 1999 (TREC-8). Available systems were able to perform at a comparable level on the TREC-6 and TREC-9 databases,

despite the order of magnitude increase in the number of items. This suggests that they should be applicable to real collections (Voorhees and Garofolo, 2000).

The Video Mail Retrieval (VMR) project at Cambridge University experimented with both word-spotting speech retrieval – recognising 35 words selected *a priori* – and indexed output from a speech recognizer to retrieve video/voice messages sent between research departments in the university and Olivetti Research Laboratory. (Young, Sparck-Jones et al., 1997; Singhal and Pereira, 1999). The video browser displaying keywords along a timeline allowed users to select regions of interest in messages to replay (Young, Sparck-Jones et al., 1997). A Multimedia Document Retrieval project has been implemented to follow on from the VMR project.

Singhal and Pereira (1999) argue that the development of very large vocabulary recognition systems resulting in fewer out-of-vocabulary (OOV) errors necessitates a review of the utility of alternative approaches. Foote (1998) described a speech recogniser with a 5,000-word dictionary as large-vocabulary, but recent developments have resulted in much larger vocabulary systems. However some words – names of people, places, products – may remain OOV and hence need retrieval via sub-word or phone lattice, as was done in the VMR project (Young, Sparck-Jones et al., 1997).

Gareth Jones, building on his involvement in the VMR project, is investigating core technology requirements to retrieve information from a mixed media collection of text, transcribed spoken word and OCR scanned documents, including the application of probabilistic models for OCR text retrieval. A system such as this would enable the BBC, for example, to amalgamate existing press cuttings archives with electronic full-text newspapers for effective retrieval (Jones, 2000).

A factor affecting the usability of retrieved spoken documents may be limitation of replay in real time. One way to overcome this is to use the capacity of hearers to understand speech at above normal rates. SpeechSkimmer is one example of a system which uses time-compression processing with algorithms to maintain original pitch and eliminate pauses to enable users to listen to spoken documents at several times normal speed (Foote, 1998).

3.3. Image Retrieval

Eakins and Graham (1999) report the growth of articles relevant to image retrieval as growing from 4 in 1991, to 12 in 1994 and 45 in 1998, from a BIDS Science Citation Index title search using ‘image*’ and ‘retriev*’. A similar search on the Web of Science for the period 1999 to July 2000 (image* AND retriev* AND (1999 OR 2000)) returned a count of 92 references. The latter search was repeated on the Citation Indexes for Arts and Humanities and Social Science returned 0 and 12 citations respectively. Rui, Huang et al. (1997; 1999) discuss present, present and future concerns and interests in image retrieval, while the report to JISC by Eakins and Graham (1999) provides an extensive and significant review of the state-of-the-art in image retrieval.

The key concern in image retrieval is the development of effective content-based image retrieval (CBIR). The parallel, of course, is with text retrieval from full-text documents. While this retrieval is imperfect, it does provide a means of access independent of human indexing (Chen and Rasmussen, 1999) which is the intention of CBIR. However, at the current state of development, the effective practical applicability of most systems tends to be limited to specialist areas and Level 1 (see

below) type queries, for example fingerprint matching, face recognition or colour matching of items in electronic mail-order catalogues (Eakins and Graham, 1999). In this discussion of image retrieval, therefore, significant concepts and findings from all areas of image retrieval are included.

3.3.1. *Content-based Image Retrieval – definitions etc.*

Content-based image retrieval (CBIR) is a technique for retrieving images on the basis of features such as colour, texture and shape automatically derived (extracted or inferred from the images themselves) without the need for human intervention (Eakins, 1996; Eakins and Graham, 1999). Retrieval by manually assigned keywords is definitely not CBIR as the term is generally understood (Eakins and Graham, 1999). CBIR is strongly related to computer science fields such as image processing and computer vision, differentiated from them by its emphasis on retrieval of desired images from a 'large' collection. . Eakins (1996) excludes from CBIR in its present form retrieval of creator, date, and/or location as primarily text retrieval problems.

3.3.2. *Research Context*

Enser (1995) noted the existence of parallel streams of research in image retrieval, divided into concept-based and content-based with little interaction between researchers and their literature. This may be explicable by the orientation of concept-based workers to large operational systems, while content-based work has focussed on solving specific problems for prototype systems often in very specific domains with small image databases (Rasmussen, 1997). This still seems to be the case with a significant interest and publication stream in computing and a librarian/information stream. There is some joint recognition of relevant work (as in Chapter 10 of Lancaster, 1998) although there appears to be a limited appreciation of IR theory from an information science perspective, and the problems of retrieving textual information. Enser (2000) argues for the continued importance of both concept-based and content-based research in image retrieval, and the development of hybrid retrieval systems. He also comments on the current paucity of shared perspectives and shared vocabularies among the researchers and workers in different domains. Rui, Huang et al. (1999) make a similar argument for the integration of disciplines.

Eakins and Graham (1999) suggest that there are over 20 large well-funded research teams in the USA, often interdisciplinary in nature – in contrast to the early 1990s when image retrieval was mainly the preserve of image processing experts. Other research activity is located primarily in Japan, Singapore, Australia, France, Holland and Germany. Most successful research groups outside USA have links there. UK research was late in the field, primarily because of limited research funding. Evaluation of CBIR effectiveness was limited in early research studies and is still difficult – though recently reported effectiveness scores are not too different from those reported for text retrieval systems (Eakins and Graham, 1999).

3.3.3. *Users – needs, searches, satisfaction*

3.3.3.1. Levels of query

Eakins (1996) proposes a typology of image queries, characterized by three levels of abstraction

- Level 1: *primitive* features such as colour, texture, shape or spatial location
- Level 2: *derived* or *logical* features such as the identity of objects shown
- Level 3: *abstract* features such as the significance of scenes depicted.

The terms *primitive* and *derived* and *abstract* are taken from Gudivada and Raghavan (1995a). They can be related to Panofsky's 1939 categorisation of

Renaissance art images as pre-iconographic (colour, shape), iconographic (describing a picture's actual subject matter) and iconologic (describing a picture's deeper artistic or religious significance). Panofsky's work is referred to by many authors, e.g. Enser (1995), Eakins (1996), and Heidorn (1999). Gudivada and Raghavan (1995) refer to Levels 2 and 3 as *semantic* image retrieval

Pre-iconographic description deals with primary or natural subject matter, and can be factual or expressional, for example a picture of a child, happiness. Iconographic analysis refers to images, stories and allegories which require social or cultural knowledge (to identify the Last Supper rather than a dinner party). Iconologic (iconographical interpretation) deals with intrinsic meaning or symbolic values (Rasmussen, 1997). Shatford (1986) differentiated Panofsky's pre-iconographic factual and expressional into objective and subjective meaning as what the picture is *Of* and what the picture is *About*. A similar distinction can be made at the iconographic level where *Of* refers to the 'specific appellation of the objects or events' and *About* refers to the 'mythical beings, ... of symbolic meanings and abstract concepts'

Chen and Rasmussen (1999) suggest that at the pre-iconographic level, image 'Ofness' and 'Aboutness' can be determined by everyday experience. At the iconographic level interpretation requires some cultural knowledge of themes or concepts (not "a sailor" but "Ulysses"). The iconologic requires interpretation at a sophisticated level using world and cultural knowledge plus a deeper understanding of the history and background of the work. They quote Svenonius (1994) that indexing aboutness at the iconographic level is equally problematic since what is symbolized is not always evident, nor is there always a simple referent to it. There is an assumption in these discussions of a shared cultural background, even at the common-sense level.

As far as CBIR is concerned, Level 1 features are objective and directly derivable from the images themselves, without the need for external knowledge. Eakins sub-divides this Level into two, retrieval of entities of a given type (e.g. a dog) and retrieval of individual entities (e.g. a picture of a Crufts' winner). Searching at Level 2 requires reference to knowledge external to the image, but is still reasonably objective. A large number of the queries noted by Enser and McGregor (1993) could be categorised as Level 2. Eakins suggests (1996) that what we identify as Level 2.1 – retrieval of objects of a given type – is equivalent to their 'non-unique' or 'non-unique refined' and Level 2.2 – retrieval of specific individual objects or persons - to their 'unique' and 'unique refined'. That is, one dimension is generality/ uniqueness, and the other is refinement by, for example, time or location. Level 3 retrieval involves significant reference to external social knowledge and high-level reasoning (e.g. to find any image of an emotional state). Eakins and Graham (1999) sub-divide Level 3 into retrieval of named entities (e.g. pictures of Scottish folk dancing) and retrieval of images with emotional content (e.g. 'suffering'). They suggest that there is no equivalent to Level 1 in text searching. Level 1 is amenable to automated indexing / retrieval while there is some evidence that some types of Level 2 searching can be. However Level 3 retrieval (e.g. find a picture of the Jarrow march) is so dependent on external world knowledge that it is difficult to see how this could ever be automated (Eakins, 1996). The approach suggested in this feasibility study, utilising synchronisation parameters, might provide a method to resolve this problem, by linking explicit references from a text file, for example, to relevant images. While CBIR systems currently operate effectively only at the lowest level, most users demand higher levels of retrieval (Eakins and Graham, 1999). Forsyth (1999) comments on the inadequacies of current technical applications for solving high level queries.

3.3.3.2. Studies of users

Little is known about the information needs of image users, the components of relevance judgements and techniques for measuring retrieval performance (Rasmussen, 1997). The view that little work has yet been published on the way users search for and use images is supported by Eakins and Graham (1999). Chen and Rasmussen suggest that The Hulton Study (Enser and McGregor, 1993; Armitage and Enser, 1997) remains the only study of its scale to examine information needs in a nondomain-specific (general subject) environment.

There are some recent studies of particular groups of users of image collections including: Markey (1988) on art history users; Ornager (1997) who proposes 5-fold typology based on type of query and role of intermediary, based on a study of journalists using newspaper archives; and again on journalists, Markkula and Sormunen (1998). Keister (1994) suggests that how users formulate queries is related to their conceptual frameworks, for example users with an artistic background frame queries in graphic terms, while health professionals are more likely to use terminology consistent with existing retrieval systems. The different roles of intermediary can also affect the way in which users interact with the retrieval process (Eakins and Graham, 1999).

Rasmussen (1997) cites the distinction made by Jose and Harper (1995) of the two main types of image retrieval activity as looking for new items (information discovery) and looking for items viewed earlier (information recovery). Searches of the former type may be related to Eakins and Graham's findings (1999) that users were not always able to express their information needs adequately, sometimes because they were not sure what they were looking for.

Users of commercial systems typically submit a keyword query which retrieves a set of images, each of which can be used to initiate a similarity search, retrieving additional images which may have been indexed under different keywords. A promising alternative appears to be using CBIR with latent semantic indexing for the retrieval of images on the Web (Eakins and Graham, 1999). Latent semantic indexing is a technique for identifying similar documents from the concepts which they contain but which are not necessarily expressed verbally (Berry, 1996; Gordon and Dumais, 1998).

3.3.4. Indexing

The normal technique for image storage and retrieval is to assign descriptive metadata in the form of keywords, subject headings or classification codes, and to use these descriptors as retrieval keys (Eakins and Graham, 1999). There are, as with text, different needs and approaches for descriptive and subject indexing.

Sarasan (1984) provides a still useful review of the principles of 'traditional' methods, dependent on significant input by human indexers, of providing subject access to image collections. Rasmussen (1997, condensed and updated as Chen and Rasmussen, 1999) provides a state of the art review of indexing images, excluding technological aspects. She distinguishes 'concept-based indexing', in which images and the objects therein are manually identified and described in terms of what they are about and what they represent, and 'content-based indexing', in which features of images (e.g. colour) are automatically identified and extracted. Concept-based indexing may be by controlled vocabulary, natural language, perhaps the use of image captions, or visual thesaurus individually or in combination. Jørgensen (1998) identified four perceptual classes as minima for image indexing – objects (the largest set in the study), people, colour and location.

Current indexing still relies largely on text descriptors or classification codes (Eakins and Graham, 1999). A review of non-text information retrieval methods for the EC (Lutes, Kutschekmanesh et al., 1996?) noted that at that time all current image archives used text-based indexing which, while not adequate for non-text, seemed generally to satisfy users. However, Eakins and Graham (1999) question this on a number of counts and suggest, from a survey of art librarians, that one reason why users were not always able to express their information needs adequately was because they did not know how to use text-based indexes.

Examples of image indexing schemes include the Art and Architecture Thesaurus, originating at Rensselaer Polytechnic, and developed by the Getty Institute which consists of approximately 120,000 terms in 7 main and 33 sub-facets (Petersen, 1994; see also http://www.ahip.getty.edu/aat_browser/); the ICONCLASS classification scheme for art images (Gordon, 1990: see also <http://iconclass.let.uu.nl/>). The ADAM (Art, Design, Architecture and Media) gateway uses the Art and Architecture Thesaurus to organise electronic resources (<http://adam.ac.uk/>).

Two image database projects, ELISE (Black and Eyre, 1995) and Déjà Vu (Gordon and Domeshek, 1998), used controlled vocabulary. In Déjà Vu users can browse through subject terms which are grouped on commonsense knowledge of library users. In response to the search term the system also indicates Broader Terms (BTs), Narrower Terms (NTs) and Related Terms (RTs) and other information with retrieved items (Chen and Rasmussen, 1999).

There have been various experiments in natural language indexing for image retrieval, some using user input. Turner (1995) and Quintana (1997) used different forms of user involvement in experiments to improve indexing effectiveness. User input to the indexing process is generally regarded as productive.

VisualSEEk allows users to add descriptors to an entire set of similar images in a single operation. VisualSEEk is a content-based image query system that allows for retrieval based on properties (such as colour, shape) of regions within an image as well as spatial locations and relationships. It provides tools for sketching query regions, identifying desired properties for regions and forming queries involving both content and spatial properties (Rasmussen, 1997).

Smeaton and Quigley (1996) used the WordNet knowledge base to compute word-word similarities in retrieval tests on manually indexed commercially available images. Dunlop and van Rijsbergen (1993) used information in hypermedia textual nodes connected with images to generate a descriptor for the image (Rasmussen, 1997).

The use of hyperlinks in indexing has been shown to be a promising area for exploitation. This has been the focus of the Multimedia Research Group at Southampton in the development of the Microcosm System and succeeding versions of MAVIS (Microcosm Architecture for Video, Image and Sound) and MAVIS 2 (Lewis, Davis et al., 1996; 1997; Dobie, Tansley et al., 1999a; 1999b). Microcosm is a system to create link databases for open hypermedia for content-based retrieval and content-based navigation. Once a link between a source selection and a destination point has been authored as a generic link it may be followed from every occurrence of the source selection in any document which has access to the link database. Thus a generic link from the word *Amsterdam* to a relevant video may be followed from any occurrence of the keyword (Lewis, Davis et al., 1997). Access to

the link database is enhanced by the use of a thesaurus, including, in concept at least, Broader Terms, Narrower Terms and automatic switching from natural language input to a controlled vocabulary (Lewis, Davis et al., 1997). The later versions of the system allow Query-by-Example as well as use of the text thesaurus. The thesaurus can be extended to include different multimedia representations with each term, for example to allow different types of *chair* or different rotations to be retrieved (Dobie, Tansley et al., 1999b).

Concept-based indexing has the advantage of providing a higher-level analysis of the image but is expensive to implement and suffers from a lack of interindexer consistency due to the subjective nature of image interpretation (Chen and Rasmussen, 1999). Estimates for manual indexing time per image (admittedly of different types) range from 7 minutes to over 40. Interindexer inconsistencies arise because of varying interpretations at the iconographic and iconologic levels. The problem is not so much that a picture is worth a thousand words as the fact that those words vary from one person to another (Keister, 1994). Eakins and Graham (1999) suggest that concrete image terms should be significant in indexing, as likely to have good inter-indexer consistency. It is easier to determine a picture's content than to interpret what is about. Krause (1988) distinguishes between 'hard' indexing (the description of what an indexer can see in the frame), and 'soft' indexing ("aboutness", the image as stimulus) (Chen and Rasmussen, 1999). However, as Chen and Rasmussen (1999) point out, interindexer consistency is problematic even in text. It is much easier to index an image for a collection with some specific use than one for use by a heterogeneous group (Shatford, 1986).

There are also problems of 'translatability' between textual and visual indexing / retrieval languages (Rasmussen, 1997, citing Svenonius, 1994). Similar points about using text to index nontextual media are made by Enser (1995) and Heidorn (1999). Heidorn uses as an example the availability of words to describe the colour of the sky at sunset.

A visual thesaurus uses a pictorial representation of the concepts in an indexing vocabulary as a direct means of access to the images, without the need to translate to and from the linguistic representation (Rasmussen, 1997). The NASA Visual Thesaurus, which is fundamentally a textual thesaurus with images associated, is an example of one form of visual thesaurus. Work on a prototype system to facilitate access to the many terabytes of data collected by NASA's Earth Observing System is reported by Alba-Flores, Koutsougeras et al. (1999).

Content-based indexing is relatively inexpensive to implement but provides a relatively low-level of interpretation of the image except in fairly narrow and applied domains. To date, very little is known about the usefulness of the access provided by content-based systems, or about user needs, satisfaction etc. (Chen and Rasmussen, 1999).

Even picture captions, which may be used in a content-based system, assume and depend on real world knowledge to make sense. Chen and Rasmussen (1999) use as an example of this the idea of a photograph of "President and Mrs. Clinton dancing" which would not be captioned "President Clinton (left) dancing with Hilary Clinton (right) at the Inaugural Ball".

Evaluating the effectiveness of CBIR versus manual indexing is not easy, because the systems are designed to answer different questions. CBIR is quicker, cheaper for some and definitely more effective when an image is difficult to describe in natural language, e.g. trademarks. However CBIR cannot cope with derived or logical

(iconographic) feature queries. Combination (hybrid) techniques are probably more effective, that is using keywords for (general) semantic content and CBIR for particular shapes, for example (Eakins and Graham, 1999). Sutcliffe, Hare et al (1997) suggest that there is little understanding of when CBIR may be profitably used in combination with, or instead of, more traditional query languages. Enser (2000) emphasises the continuing importance of verbal transactions in image retrieval.

3.3.5. *Non-verbal interfaces*

Query-by-example (QBE) allows users to submit a sample of the kind of output desired. This does, of course, depend on users having an example to submit (Eakins and Graham, 1999). Alternative QBE input formats include selecting one or more colours from a palette; textural queries, by a variety of methods; selecting a range of given shapes, using for example general shapes, boundary features and skeletal representations; sketching; spatial location. Most current research is two-dimensional, although work going on on 3-D images. Combining forms of query has been shown to be useful. Retrieval success by spatial location on own tends to be limited, but is noticeably more effective when combined with other cues, such as colour (Eakins and Graham, 1999).

The use of relevance feedback improves retrieval effectiveness and has been used successfully in several CBIR systems. One non-trivial problem concerns methods for users to convey individual notions of image similarity (Eakins and Graham, 1999). Another is the facility for different users to use different style of interaction. Interfaces for moving images, video querying and retrieval, are more complex than for still images (Eakins and Graham, 1999).

3.3.6. *Search process*

CBIR searching differs from 'traditional' IR from text databases. The aim of the latter is characterised by Eakins and Graham as the attempt to partition the document set into relevant and non-relevant items by comparison of query descriptors and inverted indexes containing multiple descriptors for each document. Images in image databases typically contain fixed-length real-valued multi-component feature vectors and the search process is intended to calculate the similarity between feature vectors for a query and stored images, essentially a similarity rank-order sorting process. Similarity clustering of images into relevance hierarchies appears to be a promising method of retrieval with the added bonus of facilitating browsing (Eakins and Graham, 1999). This does seem to downgrade the significance now attached to relevance ranking in many areas of text retrieval. As Rasmussen (1997) indicates, browsing of retrieved documents is an important process in refining a query or selecting relevant items both text retrieval systems and image retrieval systems.

CBIR operates by identifying N stored images most closely matching the query from computation of content characteristics for stored image and query. Semantic features such as the type of object present are harder to extract, although an active research topic (Eakins and Graham, 1999).

Heidorn (1999) argues that all visual information retrieval research, from the computational complexity of edge detection to national standards for museum indexing of graphical materials, is an attempt to bring the indexing model and the user's mental model into line. The computer must model the image in a way that is isomorphic (but not identical) to the human model of the image.

There are two types of correspondence that must exist between people and an image retrieval system – mental-model-to-index and cognitive-model-to-interface. The former is the extent to which a particular indexing facet is in harmony with the

cognitive/perceptual models and predispositions of the searcher. The latter is the degree of agreement between the searcher's cognitive/perceptual models and the ability to express these in the interface (Heidorn, 1999, based on Borgman, 1986). Rasmussen (1997) suggests image retrieval involves a strong component of recognition, e.g. by presentation of thumbnail images as a form of relevance feedback. Image presentation, therefore, is significant as well as image quality.

Aigrain, Zhang et al. (1996) discuss the main principles of automatic image similarity matching. Idris and Panchanathan (1997) review CBIR technology.

3.3.7. *Search engines for image retrieval*

Commercial systems include: IBM's QBIC; Virage's VIR Image Engine (used by Alta Vista AV Photo Finder); and Excalibur's Image RetrievalWare (used by Yahoo! Image Surfer). Both Alta Vista's and Yahoo!'s systems are, obviously, more limited than experimental systems on which they are based. Islip Media Inc's Mediakey Digital Video Library system is based on the Infromedia project, which is described below.

Demonstration systems include: MIT's Photobook, the face recognition technology incorporated in FaceID; Columbia University's WebSEEk; the University of Illinois's MARS (Multimedia Analysis and Retrieval System) project; and Carnegie Mellon University's Infromedia project. INRIA's Surfimage system uses multiple types of image features which can be combined in different ways and offers sophisticated relevance feedback facilities.

WebSeer was an 'early' system to facilitate image retrieval on the Web. WebSeer used image content in addition to associated text, that is the information in the image header and text surrounding it, to index images, presenting the user with a selection that potentially fitted her needs. Text cues included: image file name; image captions; ALT=Text fields (the alternate text presented if an image failed to load); HTML titles (displayed in history lists, for example); hyperlinks; and other text. Features of the image, such as colour, size, file type and size, were also used to support the retrieval process. Analysis of images and text was done off-line in the process of creating the database. At the retrieval stage the searcher could assign weights to desired characteristics (Frankel, Swain et al., 1996).

WebSEEk and ImageRover use remote engines to seek out and index images which can then be searched by keyword or content similarity, with relevance feedback (Eakins and Graham, 1999). The MetaSEEk project demonstrated the feasibility of implementing a meta-search engine on the Web (Benitez, Beigi et al., 1998). MetaSEEk was designed to select from multiple Web image search engines and interface with appropriate ones according to their performance in resolving different classes of user queries. Relevance feedback was integrated in refining the ranking process. A demonstration is available at <http://www.ctr.columbia.edu/MetaSEEk/>, with help text at <http://mahler.ctr.columbia.edu:8080/MetaSEEk/>.

A number of commercial systems for image data management include IR capabilities to varying degrees; iBase; Index+; Digital Catalogue; Fastfoto; FotoWare; Signpost; Cumulus (Eakins and Graham, 1999). Examples of users of such systems include Wellcome Trust, British Tourist Authority, British Library, Daily Express, Sky TV, North Yorkshire Police, and SCRAN.

Image retrieval using sophisticated computational techniques, such as wavelet transforms and computing differential invariants based on Gaussian derivatives at

multiple scales of images, has been reported to be effective (Eakins and Graham, 1999).

Automatic image interpretation is possible. Principal approaches seem to be by scene recognition (e.g. IRIS in 1995 could derive likely interpretations which could be used to produce text descriptors) or by object recognition (as an area within computer vision) (Eakins and Graham, 1999). Systems may learn from user defined semantic labels for defined areas, and applying these labels to areas with similar characteristics. The concept of the *semantic visual template* expresses the method of deriving labels by asking users to define query in terms of shape, colour, motion etc. and refining this by relevance feedback until the user is satisfied. The query, and resulting identified image, is given a label, such as sunset, which can be re-used and contribute to the development of a *visual thesaurus*, linking semantic concepts to a range of primitive features most likely to be relevant (Chang, Chen et al., 1998).

Forsyth (1999) describes a system which is partway between content-based and concept-based approaches, using 'detectors' of different kinds (e.g. for human bodies, trees and sunsets) which uses a set of low-level image properties to infer the existence of objects. He reports the results of an experiment in retrieving image of horses as approximately 11% recall with 67% precision (Forsyth and Fleck, 1997). He also comments on the advantage of allowing the use of 'negative examples' in improving retrieval (De Bonet and Viola, 1998) and the success in 'Blobworld' experiments of retrieving images of tigers and cheetahs (Carson, 1999; Carson, Thomas et al., 1999).

3.3.8. Information retrieval from videos

CBIR techniques are used to break up long videos into individual shots, extract the still *keyframes* summarizing the content of each shot, and search for video clips containing specified types of movement (Eakins and Graham, 1999). All but the shortest videos are made up of a number of distinct *scenes*, each of which can be broken down into individual *shots* depicting a single view, conversation or action. A common way of organizing a video for retrieval is to prepare a *storyboard* of annotated still images (often known as *keyframes*) representing each scene. Another is to prepare a series of short video clips - a process sometimes described as *video skimming*. Video queries can be categorized in Levels as still images, with the addition of queries including motion; at the lowest level, objects moving in a low parabola might be used to identify a game of tennis. Videos offer additional access points for retrieval, in that speech, music, textual credits and possibly *closed-caption text*, may be searchable either individually or in combination.

Different approaches seem to have produced similar solutions, for example dividing video sequences into different shots to select keyframes for a storyboard, which can be manually edited or accessed by CBIR techniques. A real-world application of CBIR is to use the technology (e.g. systems Virage's Videologger and Excalibur's Screening Room) to produce storyboards automatically, which can then be manually indexed for retrieval. Islip Media Inc's Mediakey Digital Video Library System and the Israel-based Media Access Technology Ltd's Visionary are currently leading edge examples of systems allowing a higher level of automation in the indexing process. The Informedia project at Carnegie-Mellon University (Wactlar, Kanade et al., 1996) has demonstrated the effectiveness of combining video with other retrieval cues and is reported to have achieved 100% recall in certain test combinations (Eakins and Graham, 1999).

The Informedia Digital Video Library project has been described as a successful pioneering system – a landmark in video retrieval - for indexing and retrieval which

integrates video, audio and other sources of information (Ponceleon, Srinivasan et al., 1998). Wactlar, Kanade et al. (1996) describe the basis of the Informedia Project as a highly accurate, speaker independent speech recognizer which is used to create transcripts of video soundtracks in a full-text information retrieval system, together with associated annotations, including closed-captioning, and credits. The digital videos are segmented into 'paragraphs' from the analysis of such information as change in camera shot, speaker changes, background music etc., and video skims produced at varying compression rates to ensure the integrity of any retrieved segment. Significant words and images from paragraphs are extracted to produce video skims which allow effective searching and browsing. The speech recognition system is explained in some detail, including recall and precision evaluations, by Witbrock and Hauptmann (1998). Christel and Pendyala (1996) discuss the contribution of trial users to the development of the Informedia interface. Sato, Kanade et al. (1998) report a 70% success rate in for word recognition from optical character recognition of digital video news, based on on-screen text and closed-captioning. One of the points they make is the significance on text captions for identifying, for example, key speakers by name and title/status.

The Multimedia Information Retrieval (MMIR) group at Dublin City University, led by Alan Smeaton, has as a key focus for the period 1997-2000 research into techniques to index and browse digital video

(<http://www.lorca.compapp.dcu.ie/~asmeaton/research.html> and elsewhere).

Primary work has been on developing shot boundary detection and representative frame selection techniques to index encoded video automatically facilitating end-user navigation and browsing (Lee, Smeaton et al., 2000). Their experimental Fischlár system is being used as a demonstrator with about 100 users who are able to record digitally and subsequently browse or replay television broadcast programmes via a Web interface.

An alternative to the use of simple keyframes is the use of mosaicing technique by the Center for Computing Technology at Bremen with IBM's ImageMiner system. Mosaicing combines several single, related shots automatically extracted from a video to be presented as a composite image. The mosaiced image is amenable to object recognition and automatic indexing to a textual description. ImageMiner has been extended to include visual queries as well as text. An illustration of this work is the retrieval of 'mountainlake' scenes (Alsuth, Hermes et al., 1997; Herzog, Miene et al., 1998).

The Digital News Project at Columbia University is intended to develop a suite of effective interoperable tools with which people can find relevant information from text, images, audio and video in videos from distributed sources (Aho et al., 1998). A key intention is to enable users to keep track of news in specific areas. The WebClip system facilitates searching and browsing of video in compressed format over the Web, using visual features with other multimedia features and manual indexing (Chang, Smith et al., 1997). They stress the limited suitability of 'current' compression standards (JPEG, MPEG-1, MPEG-2) for dynamic feature extraction and the current lack of satisfactory methods for measuring the effectiveness of image search techniques although MPEG-7 will incorporate the evaluation process as a mandatory part of the standard development process (Chang, Smith et al., 1997; Chang, Huang et al. 1999). These authors suggest that MPEG-7 will facilitate searching, by Query-by-example (QBE), by sketch and by using MPEG-7 multimedia descriptions. Their work includes the use of a number of semantic video indexing techniques, including Semantic Templates and Semantic Visual Templates, where relevance feedback allows the system to learn what sorts of images are likely to be relevant to a query on 'sunsets' or 'high jumpers'. Their suggestion that weights

assigned to features for retrieval may be thought of as the user's belief in the relevance of the feature with respect to the object to which it is attached **may** indicate the value of using an approach such as Dempster-Shafer evidence combination metrics for retrieval as Jose and Harper (1997). The concept of using SVTs appears to work well in practice, boosting the recall of 'sunsets' from a large heterogeneous database to 50% from the 10% retrieved by the basic query. The key focus of the Digital News Project is on retrieval from television news broadcasts and includes, for example, experiments in retrieving reports of downhill skiing and El Niño.

Ponceleon, Srinivasan et al. (1998) report on experiments at the IBM Almaden Research Center to retrieve information from video-recorded technical talks and presentations. Their semi-automated CueVideo system integrates voice and manual annotation, attachment of related data, visual search technologies (IBM's QBIC) and novel storyboard generation to provide a system where the user can incorporate the type of semantic information that automatic techniques fail to obtain. Their system can represent several minutes of video on a single HTML page, a compression ratio of c. 100:1. While the use of a speech recognition interface, with controlled and uncontrolled vocabulary input, allows untrained personnel to create the metadata associated with a video, they emphasise the relatively expensive but cost-effective role of knowledgeable domain experts to contribute to the cataloguing process.

Foote, Boreczky et al. (1998) discuss work on information retrieval from video at the FX Palo Alto Laboratory, again using recordings of meetings as the document type. Their system generates confidence scores for points of interest, for example combining detection of shot changes with periods of silence to identify change of speaker. The retrieval interface is via a browser which allows users to determine confidence thresholds and to fast-forward through non-significant sections.

Smith (1999) presents an overview of access to digital video libraries via the Internet within the context on the Next Generation Internet Initiative.

3.4. Metadata, standards

Standards development can be identified as image compression, query specification and metadata description. By far the most important emerging standard is MPEG-7, which will define search features for all kinds of both still image and video data (Eakins and Graham, 1999).

Eakins and Graham suggest that until a specific multimedia standard for data description and representation (i.e. MPEG-7) is available, key standards of interest are image compression (JPEG, MPEG-2); query specification (e.g. SQL type languages) and metadata standards such as RDF (Resource Description Framework) and XML. Feature extraction from compressed data, especially MPEG-compressed video, offers significant time benefits as well as supporting the use of motion vector information. Query specification is seen to be important because CBIR search arguments (arrays of real numbers representing extracted image features, and specifications of similarity matching algorithms) are not easily handled in, for example, current Z39.50 applications (Eakins and Graham, 1999). Current metadata standards, e.g. Dublin Core, only provide for an image DTD and the use of keywords to describe content. The more generic RDF potentially affords much greater relevance and hospitality to CBIR. MPEG-7 is seen as the key standard to represent multimedia content, including low-level descriptions (Level 1 type) and high-level abstractions (Levels 2 and 3). While the primary orientation of MPEG-7 is towards digitized video it should be entirely hospitable to still images. However, there are

several concerns about the development of MPEG-7, not least that because the field is immature the 'best' features / methods may not be selected for inclusion since they cannot yet be identified, and semantic retrieval cues will still have to be added manually (Eakins and Graham, 1999).

Dempsey and Heery (1998) discuss the different meanings of 'metadata', the range of potential uses and the variety of metadata formats for particular user communities and applications, presenting a number of typologies of metadata formats and associated technologies. Martin, Powley et al. (2000) and Vieira, Biajiz et al. (1999) present accounts of experimental research using metadata for information retrieval. The latter have implemented a Multimedia Object Server which supports content-based searching of image, text and video. MHEG-5 metadata defined by the author are used for fuzzy set searching while exact match searching of audio seems to be primarily by indexing transcripts and thesaurus creation. Martin, Powley et al. (2000) describe the extraction and use of metadata from data sources to create a metadatabase, which may include hyperlinks, to support searching. One example they use is to 'Find all documents related to Unit 4 of the distributed Systems course taught by Prof. Martin'. Metadata is defined for their purposes as descriptions of the properties of, and the relationships present in, the data and data sources.

A significant new metadata project is Harmony, funded by JISC, DTSC and the NSF, focussed on the multimedia resources in digital libraries. Key issues which Harmony will address are: collaborative refinement of metadata standards; investigating a conceptual model of interoperability, mechanisms for expressing such a conceptual model; and developing mechanism to map between community specific vocabularies using such a conceptual model (Harmony, 1999). A first attempt at a *Logical model for metadata interoperability* was put forward for discussion at workshop in January 2000 (Brickley, Hunter et al., 1999).

Many of the best sources of information about metadata are on the Web (Dempsey and Heery, 1998). As part of the Synchronisation Project an annotated list of metadata sources has been created and made available, with links, at: <http://www.mmu.ac.uk/cerlim/projects/synchro/synchro.htm>. A particular focus of the site is the proposed SMIL metadata standard. A sample demonstration of SMIL-based presentation is accessible through the site.

3.5. Combined approaches and the use of synchronisation

Srihari describes a system which combines speech recognition, natural language processing and image understanding to identify and label areas, roads and buildings in an image (1997). Srihari and Zhang (1999) describe experiments combining the use of text processing of picture captions with content-based image retrieval, in both the indexing and retrieval phases, although the work reported used text as the retrieval cue with image filtering only to enhance precision.

Wu, Miller et al. (1998) argue that multimedia presentation models, as well as facilitating the retrieval and playback of presentations, enable the browsing and searching of presentation libraries to support the separate use of presentation components. They give as examples an author who wishes to retrieve all (other) objects synchronised with some specific target object. They envisage a multimedia presentation and authoring system which should provide an integrated, flexible composition and query capability. While a primary retrieval capability of such a system is to facilitate retrieval by temporal synchronisation characteristics it should also permit content and attribute based queries.

Hürst and Müller (1999) suggest that in their synchronisation model for recorded presentations, the use of time stamps across all media streams, primarily related to presentation slide changes, supports retrieval by facilitating browsing and navigation with random visible scrolling. The links between media streams allow the selection of one media stream to be selected for indexing and the retrieval of other objects by their links. While the model is still under development elements of it have been implemented in a multi-university project (VIROR – <http://www.viror.de>) as a step in the development of a virtual university.

Yoshitaka and Ichikawa (1999) discuss approaches to content-based retrieval for multimedia databases, among which they include individual nontextual/numeric media, primarily from an implicit standpoint of DBMS. One of their motivating factors was the significance of temporal (synchronization) and spatial relations in retrieving multimedia data, for the management of which database systems based on relational data models provide only limited facilities. Lee, Sheng et al. (1999) discuss three presentation graph languages for querying presentation graphs using temporal operators and for querying presentation graphs for content information.

4. Experimental Design

The overall design for a practical SOR experiment has involved consideration of the state of the art and the various issues outlined in section 2 related to the SMIL standard itself. The following stages are required:

4.1. Establishment of test collection

It will be necessary to establish a reasonably sized test collection of SMIL-compliant multimedia objects, showing a degree of heterogeneity in structure and micro-object composition. We are not aware of the existence of a suitable collection of SMIL-compliant objects at the present time.

It will therefore be necessary to acquire or create suitable presentations for inclusion in a collection and then to edit the presentations to provide suitable indexing. Initial discussions (e.g. with SCRAN) reveal a willingness of multimedia collection holders to become involved, but a lack of SMIL compliance and problems in relatively simple make-up of objects – for example, most multimedia packages incorporate audio, still images and video but use limited synchronicity (normally limited to one audio and one image at a time) – will need to be resolved. In order to test the SOR approach thoroughly in experimental conditions, a carefully designed collection would be needed. In other words, it would be essential to put effort into ensuring that the collection displayed examples of all the different elements on which retrieval could be based. Not only would comprehensive metadata (at the object and micro-object levels) be needed, but care would need to be taken to ensure that the different media used contained retrievable content, the attributes of which were known, in order that testing of software and algorithms could take place.

Because of the complexities inherent in the SOR approach, we would suggest that a test collection should be domain limited initially. For example, the use of a domain such as, say, the music of a specific composer or a specific locality's history and culture would avoid the worst trans-domain semantic and other problems. The development of a test collection would be a challenging piece of work in its own right, but would be an essential first step to the establishment of an experimental

programme. It would be helpful to consult with the TREC team in defining this collection.

It is perhaps worth noting here that although the test collection would need to be carefully crafted, this does not mean that it would have to be entirely artificial. There would be merit in identifying a nascent, real-life collection which could be SMIL-enabled, not least because experimental work could then contribute to the solution of real-world problems and real-world queries could inform the evaluation of retrieval.

4.2. User input

The analysis of the user query into mono-media components will itself be a complex process. Although in theory this could be automated, so that software converted a query expressed, let us say, as a text string ("*Find something about the sun*") into a series of media specific queries ("*sun*" [text]; "*sun*" [audio wave file, retrieved from a 'thesaurus']; "*sun*" [image description, "*round yellow object on blue background*", retrieved from another 'thesaurus'; and so on), in practice this is currently impractical (or at least a different project!). In practice it would be feasible to generate a multi-media query by providing the user with an interface which enabled descriptions to be selected/entered in a series of 'channels' e.g. text box, select from image thumbnails to find 'things like', audio input terms (the problem of segmenting and recognising audio-based queries is, however, non-trivial – and we noted in section 3 that non-verbal sounds pose even greater problems than speech).

It would be possible to make user input an iterative process by presenting to the user examples of retrieval terms, especially where non-textual retrieval was being undertaken. Thus, to extend the above example, an initial user query for an image of the sun might result in a series of thumbnails (extracted by reference to a thesaurus) being displayed and the user being invited to select from them or rank them in order of relevance. This additional user input could also be used to enable the system to learn from search preferences and possibly provide personalisation of results, although again this would be beyond the scope of the initial experiment.

Fig. 1 (on the next page) shows a mock-up of the type of user interface we have in mind.

4.3. Query Analysis

The user query is analysed into query statements capable of being applied within specific mono-media contexts as discussed above. So, for example, the query might be expressed as a series of text strings (some perhaps derived from speech) for matching against a text file, as a voice/sound wave format for matching against an audio file and so on. A later project might examine how, for example, text strings might be applied to an intermediate image thesaurus to create image queries not present in the original input data. As indicated above a possible way to use such query enhancement would be to display a set of images to the user and invite selection of the nearest matches.

Enter text here

Text search



Select



Select



Select

Shape search



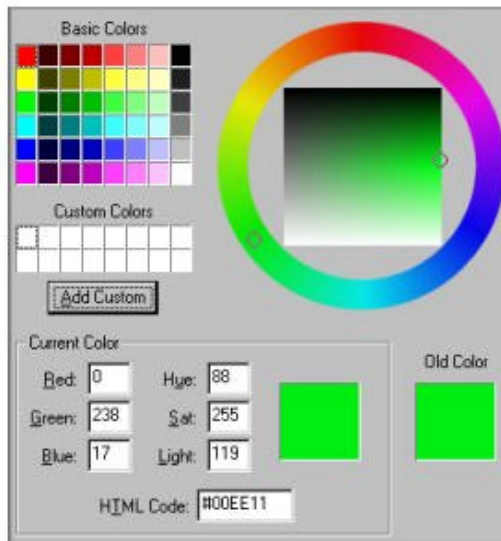
Select



Select



Select



Colour search

Sound

More

Sound search

Motion

More

Motion search

* Select any combination of the above to create your search

* Click to start

Search

Fig. 1: SOR User Input Screen

4.4. Search of packages

The first search process should consist of a search of the metadata associated with each package (as a whole). A ranking of objects in the collection can be achieved in this way, although it would be important not to discard packages with apparent zero relevance from this process, since many objects may lack metadata and much package-level metadata may be irrelevant to specific component micro-objects. (An example of this could be a package designed to illustrate a geographical area which contained images of flora and fauna of the region but described merely by geographical location – the metadata would probably produce a ‘zero relevance’ result for a query for a particular plant or animal even if they were in fact represented in micro-objects).

This process leads to a score ‘A’ being computed for each package. As with other searches, this stage is intended to reveal clues, not to stand on its own as a retrieval process. Where some packages do not have associated metadata it may be necessary to provide a neutral score (see also section 4.5 below).

4.5. Search of micro-objects’ metadata

The search of the micro-objects themselves could be carried out in a number of ways. For simplicity we suggest that the first step should be to check all micro-objects for embedded metadata, to search on this and to rank micro-objects for relevance accordingly. Again, many micro-objects will lack metadata so again all that is produced is a set of clues to relevance. This process allows a score B_x to be computed, where x is the micro-object sequential number*. To avoid micro-objects without metadata being given the lowest scores it would again be appropriate to calculate a mean score for all micro-objects which have metadata and use this to ensure that metadata-less micro-objects score neutrally. It will be necessary to identify the most appropriate scoring model for the purpose; this will form part of the research itself.

4.6. Search of micro-objects’ content

Here we reach the core of the approach. We assume here that the aim is to retrieve micro-objects, although the process of retrieving packages is similar, requiring only a method for combining the retrieval scores of constituent micro-objects. The process might take the following base model:

1. The synchronisation file is analysed to identify the first micro-object – note again that ‘first’ here is arbitrarily defined, simply implying that there is some ordering which ensures that each micro-object will be examined in sequence. The media type should be defined in the SMIL *type* attribute and this should be used to determine the type of retrieval process to be engaged. Using this process, micro-object 1 is examined and a score C_1 computed. Again, the establishment of a scoring algorithm would be an important part of the experimental research.
2. Micro-objects which are due to be played simultaneously with the first object (this should be interpreted as meaning that there is some temporal overlap, not that they start and end at the same time) are

* It should be noted that micro-objects may be regarded as being in any order for this purpose, provided only that they can be processed sequentially. Thus it may be appropriate to treat them as sequenced by filename, by file type, by ‘running order’ or in a variety of other ways.

then examined against whatever media retrieval process is appropriate. The results of each analysis are used to modify C1, producing D1: again we have not defined the exact algorithm to be used, although it may be noted that different media types might be weighted differently if it was found that, say, text retrieval produced more reliable results than content-based image retrieval. The effect of this stage is that C1 is increased whenever a high probability of a match in a micro-object due for simultaneous play is identified.

3. When micro-object 1 and all other micro-objects with which it shares some temporal space have been processed, the next micro-object in the sequence is examined and steps 1 & 2 repeated, providing scores D2, D3, D4 Dn. It should be noted that because the synchrony defined for each micro-object, it will be necessary to use the relevance score for each multiple times – clearly it will be more efficient to store these and re-use them during the process than to recompute them.
4. At the end of the object analysis a relevance score for each micro-object is available. This is combined (again, experimentation will be needed to establish the algorithm to be used) with the overall score A and the metadata score Bn to produce a ‘final’ score for each micro-object, En.
5. When all objects in the collection have been analysed, the micro-object scores En can be ranked to indicate the most likely matches overall. These micro-objects can then be displayed to the user.

Where the object is to retrieve multimedia packages rather than micro-objects, clearly a score for each macro-object can be computed by combining (again with a suitable, yet to be defined, algorithm) the scores for all constituent micro-objects and then ranking these overall scores.

5. Conclusions

The feasibility study has shown that the SMIL standard provides a powerful basis for the retrieval of both micro-objects and multimedia packages. Not only does the basic SOR process suggest a way of enhancing retrieval within heterogeneous, multimedia collections, but the standard’s additional features (such as support for accessibility and language variants) provides a fertile territory for associated research.

The next stages of this research programme will involve the development of a suitable test collection of SMIL-compliant material and the establishment of a major experimental programme to enable the ideas explored in outline in the feasibility study to be subjected to in-depth research and development. It is the intention of the authors to pursue funding to enable these programmes of research to be established as soon as possible.

It is clear from our research that SOR has enormous potential for application in networked information environments containing heterogeneous collections of multimedia objects. As such environments (which, after all, include the World Wide Web) proliferate, as the number of multimedia packages mushrooms and as the cost of computing power reduces, it is to be expected that SOR approaches will prove ever more valuable.

6. References

Note: Many of the references included in this bibliography are available in electronic format, either for a fee (subscription and/or pay per view services) or without charge. Only electronic sources which are freely available are cited here. Sources were accessed during the period May to August 2000.

Aho, A., Chang, S.-F., et al. (1997) Columbia Digital News System: an environment for briefing and search over multimedia information *4th International Forum On Research and Technology Advances in Digital Libraries (ADL '97)*, 13p. (PDF from: <http://citeseer.nj.nec.com/pdf/31812>)

Aigrain, P., Zhang, HongJ., et al. (1996) Content-based representation and retrieval of visual media - a state-of-the-art review *Multimedia Tools and Applications*, 3(3), 179-202. (PDF from: <http://citeseer.nj.nec.com/aigrain96contentbased.html>)

Alba-Flores, R., Koutsougeras, C., et al. (1999) Content-based search prototype for image databases. (Meta-data '99: Third IEEE Meta-data Conference, April 6-7, Bethesda) (URL: <http://www.computer.org/proceedings/meta/1999/papers/44/ralba.html>)

Alsuth, P., Hermes, T., et al. (1997) *ImageMiner: Intelligent Retrieval for Images and Video*. Bremen: University of Bremen. (Technischer Bericht nr. 1, 1997. Postscript from: <http://www.tzi.de/tzi/berichte/kurzfassung001.html>)

Alsuth, P., Hermes, T., et al. (1998) On video retrieval: content analysis by ImageMiner, 236-49 in: *Storage and Retrieval for Image and Video Databases VI*, 28-30 January 1998, San Jose: SPIE Proceedings 3312, edited by I.K. Sethi and R. Jain.

Ardizzone, E. and La Cascia, M. (1997) Automatic video database indexing and retrieval *Multimedia Tools and Applications*, 4, 29-56.

Armitage, L.H. and Enser, P.G.B. (1997) Analysis of user needs in image archives *Journal of Information Science*, 23(4), 287-99.

Arms, W.Y. (2000) Automated digital libraries: how effectively can computers be used for the skilled tasks of professional librarianship? *D-Lib Magazine*, 6(7/8), 11p. (URL: <http://webdoc.gwdg.de/edo/d-lib/dlib/july00/arms/07arms.html>)

Autonomy (2000) Autonomy acquires acclaimed speech recognition company SoftSound. (Press release, 18 May 2000) (URL: <http://www.autonomy.com/press/prsoftsound.html>)

Bainbridge, D., Nevill-Manning, C.C., et al. (1999) Towards a digital library of popular music, 161-69 in: *Digital Libraries '99: The Fourth ACM Conference On Digital Libraries, Berkeley, CA, August 11-14, 1999*, edited by E.A. Fox and N. Rowe. New York: ACM.

Bell, I. (1998) Images for the tourist: the experiences of the British Tourist Authority and English Tourist Board in setting up a digital image collection *Vine*, 107, 21-24.

Benitez, A.B., Beigi, M., et al. (1998) A content-based image meta-search engine using relevance feedback *IEEE Internet Computing*, 2(4), 59-69. (URL: <http://www.ctr.columbia.edu/MetaSEEK/MetaSEEK.html>)

Berry, M.W. (1996) Latent Semantic Indexing. (URL: <http://www.cs.utk.edu/~berry/lsi++/node5.html>)

Bjarnestam, A. (1998) Text-based hierarchical image classification and retrieval of stock photography, in: *The Challenge of Image Retrieval Research Workshop, Newcastle Upon Tyne, 5 February 1998*.

Black, K. and Eyre, J. (1995) The ELISE Project (electronic objects), 70-78 in: *Proceedings of the Second Electronic Library and Visual Information Research Conference, May 1995, De Montfort University (ELVIRA 2)*, edited by M. Collier and K. Arnold. Milton Keynes: De Montfort University.

Blackburn, S. and DeRoure, D. (1998) A tool for content based navigation of music, in: *ACM Multimedia 98 - Electronic Proceedings*. (URL: http://www.acm.org/sigs/sigmm/MM98/electronic_proceedings/blackburn/)

Boehm, C. (2000) Systems for music representation and retrieval. ((section 2 of MuTaTeD report)) (Word document from: <http://www.pads.ahds.ac.uk/MuTaTeD2home.html>)

Bolle, R.M., Yeo, B.L., et al. (1998) Video query: research directions *IBM Journal of Research and Development*, 42(2), 233-52.

De Bonet, J.S. and Viola, P. (1998) Structure driven image database retrieval, 866-72 in: *Advances in Neural Information Processing Systems*, Vol. 10, edited by M.I. Jordan et al. Cambridge: MIT.

Borgman, C.L. (1986) The user's mental model of an information retrieval system: an experiment on a prototype online catalog *International Journal of Man-Machine Studies*, 24, 47-64.

Brewer, E. (2000) Inktomi inside. ((Report of presentation at 5th Annual Search Engine Meeting)) (URL: wysiwyg://28/http:websearch.about...earch/library/weekly/aa041900a.html)

Brickley, D., Hunter, J., et al. (1999) A logical model for metadata interoperability. (URL: <http://www.ilrt.bris.ac.uk/discovery/harmony/docs/abc/abcdraft.html>)

Brown, P., Hilderley, R., et al. (1996) The democratic indexing of images *New Review of Hypermedia and Multimedia: Applications and Research*, 2, 107-20.

Carson, C. (1999) Welcome to Blobworld! (URL: <http://www.cs.berkeley.edu/~carson/blobworld/>)

Carson, C., Thomas, M., et al. (1999) Blobworld: a system for region-based image indexing and retrieval, 509-16 in: *Third International Conference On Visual Information Systems (Proceedings)*, edited by D.P. Huijsmans and A.W.M. Smeulders. New York: Springer.

- La Cascia, M., Sethi, S., et al. (1998) Combining textual and visual cues for content-based image retrieval on the world-wide web, 24-28 in: *IEEE Workshop On Content-Based Access of Image and Video Libraries (Proceedings)*.
- Chang, S.-F., Chen, W., et al. (1998) Semantic visual templates: linking visual features to semantics, 531-35 in: *IEEE International Conference On Image Processing (ICIP'98), October 1998, Chicago*. (PDF from: <http://citeseer.nj.nec.com/ cachedpage/173467/1 cached page>)
- Chang, S.-F., Huang, Q., et al. (1999) Multimedia search and retrieval, 28p in: *Advances in Multimedia: Systems, Standards and Networks*, edited by A. Puri and T. Chen. New York: Marcel Dekker. (PDF from: <http://www.ctr.columbia.edu/~sfchang/chapter-99.pdf>)
- Chang, S.-F., Smith, J.R., et al. (1997) Finding images/video in large archives: Columbia's content-Based Visual Query Project *D-Lib Magazine*, (February), 9p. (URL: <http://www.dlib.org/dlib/february97/columbia/02chang.html>)
- Chen, H.-L. and Rasmussen, E.M. (1999) Intellectual access to images *Library Trends*, 48(2), 291-302.
- Chiaramella, Y. (1997) Browsing and querying: two complementary approaches for Multimedia Information Retrieval, 15p. (Hypertext - Information Retrieval - Multimedia 1997, September 29 - October 2 1997, Dortmund) (URL: <http://ls1-www.cs.uni-dortmund.de/HIM97/Konferenzband/Chiaramella/paper0.htm>)
- Christel, M.G. and Pendyala, K. (1996) Informedia goes to school *D-Lib Magazine*, (September), 5p. (URL: <http://www.dlib.org/dlib/september96/informedia/09christel.html>)
- Chung, S.M. and Lee, J. (1998) Information discovery on the Internet, 146-57 in: *Encyclopedia of Librarianship and Information Science*, Vol. 62, Supplement 25, edited by A. Kent and C.M. Hall.
- Dempsey, L. and Heery, R. (1998) Metadata: a current view of practice and issues *Journal of Documentation*, 54(2), 145-72.
- Desai, B.C., Shinghal, R., et al. (1999) CINDI: a virtual library indexing and discovery system *Library Trends*, 48(1), 209-33.
- Dobie, M.R., Tansley, R.H., et al. (1999a) MAVIS 2: a new approach to content and concept based navigation, 9/1-5 in: *Proceedings of the IEE Colloquium On Multimedia Databases and MPEG-7, Institution of Electrical Engineers 99(056), January 1999*.
- Dobie, M.R., Tansley, R.H., et al. (1999b) A flexible architecture for content and concept-based multimedia information exploration, 15p in: *CIR-99: The Challenge of Image Retrieval, Newcastle Upon Tyne, February 25-26, 1999*. (URL: <http://www.mmrg.ecs.soton.ac.uk/publications/archive/dobie1999/html/>)
- Downie, J.S. (2000a) MUSIC IR 2000: International Symposium on Music Information Retrieval. (posting to ninch-announce@cni.org, 17 March 2000)
- Downie, J.S. (2000b) Access to music information: the state of the art *Bulletin of the American Society for Information Science*, 26(5), 5p. (URL: <http://www.asis.org/Bulletin/June-00/downie.html>)

- Dunlop, M.D. and van Rijsbergen, C.J. (1993) Hypermedia and free text retrieval *Information Processing and Management*, 29(3), 287-98.
- Eakins, J.P. (1996) Automatic image content retrieval, 123-35 in: *Proceedings of the Third International Conference On Electronic Library and Visual Information Research (ELVIRA3)*. Milton Keynes: De Montfort University.
- Eakins, J.P. and Graham, M.E. (1999) Content-based Image Retrieval. (Report JTAP 39) (URL: <http://www.jtap.ac.uk/reports/htm/jtap-039.html>)
- Ellis, D., Ford, N., et al. (1998) In search of the unknown user: indexing, hypertext and the World Wide Web *Journal of Documentation*, 54(1), 28-47.
- Enser, P.G.B. (1995) Pictorial information retrieval *Journal of Documentation*, 51(2), 126-70.
- Enser, P.G.B. and McGregor, C.G. (1993) *Analysis of Visual Information Retrieval Queries*. London: British Library. (BLR&DD Report no. 6104)
- Enser, P.G.B. (2000) Visual image retrieval: seeking the alliance of concept-based and content-based paradigms *Journal of Information Science*, 26(4), 199-210.
- Ewell, K. and Adi, T. (1997) The elusive electronic text. (19p) (URL: <http://www.readware.com/elusive.html>)
- Feldman, S. (1999) Search engines, 218-43 in: *Encyclopedia of Librarianship and Information Science*, Vol. 64, Supplement 27, edited by A. Kent and C.M. Hall.
- Foote, J. (1998) An overview of audio information retrieval , (PDF from: <http://citeseer.nj.nec.com/did/190917>)
- Foote, J., Boreczky, J., et al. (1998) An intelligent media browser using automatic multimodal analysis, 11p in: *ACM Multimedia 98 - Electronic Proceedings*. (URL: http://www.acm.org/sigs/sigmm/MM98/electronic_proceedings/foote/)
- Forsyth, D.A. (1999) Computer vision tools for finding images and video sequences *Library Trends*, 48(2), 326-55.
- Forsyth, D.A. and Fleck, M.M. (1997) Body plans, 678-83 in: *Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition, San Juan*. Los Alamos: IEEE Computer Society. (PDF from: <http://citeseer.nj.nec.com/pdf/64314>)
- Frankel, C., Swain, M.J., et al. (1996) *WebSeer: An Image Search Engine for the World Wide Web (Technical Report 96-14)*. Chicago: University of Chicago. (PDF from: <http://infolab.nwu.edu/webseer/>)
- Franklin, A. (1998) Image indexing in the Bodleian ballads project *Vine*, 107, 51-57.
- Furner, J. (1998) Digital images in libraries: an overview *Vine*, 107, 3-7.
- Ghias, A., Logan, J., et al. (1995) Query by humming - musical information retrieval in an audio database, 231-36 in: *Proceedings of the Third ACM International Conference On Multimedia 95: San Francisco, California, November 5-9*. (URL: <http://www2.cs.cornell.edu/zeno/Papers/humming/humming.html>)

Ghosh, H. and Chaudhury, S. (1998) Knowledge-based retrieval of multimedia documents, 9p in: *International Conference On Multimedia for Humanities, October 5-8, 1998*. (URL: <http://www.ignca.nic.in/clcnf140.htm>)

Gluck, M. (2000) The use of sound for data exploration *Bulletin of the American Society for Information Science*, 26(5), 4p. (URL: <http://www.asis.org/Bulletin/June-00/gluck.html>)

Goodrum, A. and Rasmusen, E. (2000) Sound and speech in information retrieval: an introduction *Bulletin of the American Society for Information Science*, 26(5), 2p. (URL: <http://www.asis.org/Bulletin/June-00/godrumrasmussen.html>)

Gordon, A.S. and Domeshek, E.A. (1998) Déjà Vu: a knowledge-rich interface for retrieval in digital libraries, 127-34 in: *IUI 98 (International Conference On Intelligent User Interfaces, January 6-8 1998, San Francisco)*. New York: ACM Press.

Gordon, C. (1990) An introduction to ICONOCLASS, 233-44 in: *Terminology for Museums, Proceedings of an International Conference, Cambridge, 1988*, edited by D.A. Robert.

Gordon, M.D. and Dumais, S. (1998) Using latent semantic indexing for literature based discovery *Journal of the American Society for Information Science*, 49(8), 674-85.

Greene, S., Marchionini, G., et al. (2000) Previews and overviews in digital libraries: designing surrogates to support visual information seeking *Journal of the American Society for Information Science*, 51(4), 380-93.

Gudivada, V.N. and Raghavan, V.V. (1995) Content-based image retrieval systems *IEEE Computer*, 28(9), 18-22.

Hahn, T.B. (1998) Text retrieval online: historical perspectives on Web search engines *ASIS Bulletin*, (April/May), 8p. (URL: <http://www.asis.org/Bulletin/Apr-98/hahn.html>)

Harmony (1999) About Harmony. (URL: <http://www.ilrt.bris.ac.uk/discovery/harmony/>)

Heidorn, P.B. (1999) Image retrieval as linguistic and nonlinguistic visual model matching *Library Trends*, 48(2), 303-25.

Herzog, O., Miene, A., et al. (1998) Integrated information mining for texts, images, and videos *Computers and Graphics*, 22(6), 675-85.

Hill, G.J., Wilkins, R., et al. (1993) Open and reconfigurable hypermedia systems: a filter-based model *Hypermedia*, 5(2), 103-18. (URL: <http://www.bib.ecs.soton.ac.uk/data/1422/html/html/>)

Hogan, M., Jörgensen, C., et al. (1991) The visual thesaurus in a hypermedia environment: a preliminary exploration of conceptual issues and applications, 202-21 in: *International Conference On Hypermedia and Interactivity in Museums, 1991 October 14-16, Pittsburgh*, edited by D. Bearman. Pittsburgh: Archives and Museums Informatics.

Huang, T., Mehrota, S., et al. (1997) Multimedia Analysis and Retrieval System (MARS) project, 101-17 in: *Digital Image Access and Retrieval: 33rd Annual Clinic On*

Library Applications of Data Processing, March 24-26, Urbana, edited by P.B. Heidorn and B. Sandore. Urbana-Champaign: University of Illinois.

Hürst, W. and Müller, R. (1999) A synchronization model for recorded presentations and its relevance for information retrieval, 14p in: *Multimedia '99 Conference Proceedings*. (URL: <http://www.woodworm.cs.uml.edu/~rprice/ep/huerst/>)

Idris, F. and Panchanathan, S. (1997) Review of image and video indexing techniques *Journal of Visual Communication and Image Representation*, 8, 289-310.

Jørgensen, C. (1998) Attributes of images in describing tasks *Information Processing and Management*, 34(2/3), 161-74.

Jones, G. (2000) Information retrieval for mixed-media collections. (ESPRC ref: GR/N04034) (<http://www.dcs.ex.ac.uk/~gareth/irmmc.html>)

Jose, J.M. and Harper, D.J. (1997) Epic: a photograph retrieval system based on evidence combination approach. (TR-97/2, SCMS, Robert Gordon University) (PDF from: http://www.scms.rgu.ac.uk/publications/97/97_2.shtml)

Jose, J.M. and Harper, D.J. (1995) An integrated approach to image retrieval *New Review of Document and Text Management*, 1, 167-81.

Keister, L.H. (1994) User types and queries: impact on image access systems, in: *Challenges in Indexing Electronic Text and Images*, edited by R. Fidel, T.B. Hahn, et al. Metford: ASIS/Learned Information.

Korfhage, R.R. (1997) *Information Storage and Retrieval*. New York: Wiley.

Kowalski, G. (1997) *Information Retrieval Systems: Theory and Implementation*. Boston: Kluwer Academic.

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

Lancaster, F.W. (1998) *Indexing and Abstracting in Theory and Practice* 2nd ed. Urbana-Champaign: University of Illinois.

Lancaster, F.W. (1979) *Information Retrieval Systems: Characteristics, Testing and Evaluation*. New York: Wiley-Interscience.

Large, A., Tedd, L.A., et al. (1998) *Information Seeking in the Online Age: Principles and Practice*. London: Bowker Saur.

Lee, H., Smeaton, A.F., et al. (2000) The Fischlár digital video recording, analysis and browsing system, 10p in: *Proceedings of RIAO2000, April 2000, Paris*. (PDF from: <http://www.compapp.dcu.ie/~asmeaton/pubs-list.html>)

Lee, T., Sheng, L., et al. (1999) Querying multimedia presentations based on content *IEEE Transactions of Knowledge and Data Engineering*, 11(3), 361-85.

Lewis, P.H., Davis, H.C., et al. (1996) Media-based navigation with generic links, 215-23 in: *Hypertext '96: Seventh ACM Conference On Hypertext*. (URL: <http://www.cs.unc.edu/~barman/HT96/P38/ht96.html>)

- Lewis, P.H., Davis, H.C., et al. (1997) Towards multimedia thesaurus support for media-based navigation, 111-18 in: *Image Databases and Multimedia Search*, edited by A.W.M. Smeulders and R.C. Jain. Amsterdam: World Scientific. (URL: <http://www.bib.ecs.soton.ac.uk/data/1423/html/html>)
- Li, G. (1999?) Audio retrieval based on wavelet. (URL: <http://www.eecis.udel.edu/~gli/audindexing.html>)
- Lutes, B., Kutschekmanesh, S., et al. (1996?) Study of non-textbased information retrieval - state of the art. (European Commission, Report ELPUB 106) (Word file from: <http://158.50.95:10080/ie/en/studies.html> - ELPUB106)
- Marchionini, G., Dwiggins, S., et al. (1993) Information seeking in full-text end-user-oriented search-systems: the roles of domain and search expertise *Library and Information Science Research*, 15(1), 35-69.
- Markey, K. (1988) Access to iconographical research collections *Library Trends*, 37(2), 154-74.
- Markkula, M. and Sormunen, E. (1998) Searching for photos - journalists' practices in pictorial IR, in: *The Challenge of Information Retrieval Workshop, Newcastle Upon Tyne, 5-6 February 1998 (CIR-98)*, edited by J.P. Eakins, D.J. Harper, et al.
- Martin, P., Powley, W., et al. (2000) Using metadata to query passive data sources *International Journal of Cooperative Information Systems*, 9(1 & 2), 147-69.
- Missingham, R. (1999) Perspectives on DL '99: Fourth ACM Conference on Digital Libraries *D-Lib Magazine*, (September 1999), 7p. (URL: <http://www.dlib.org/dlib/september99/09missingham.html>)
- OMRAS (1999) OMRAS (Online Music Recognition and Searching). (URL: http://www.dli2.nsf.gov/international_projects/JISC/OMRAS/index.html)
- Oppenheim, C., Morris, A., et al. (2000) The evaluation of WWW search engines *Journal of Documentation*, 56(2), 190-211.
- Ornager, S. (1997) Image retrieval: theoretical and empirical user studies on accessing information in images, 202-11 in: *ASIS '97: Proceeding So the 60th ASIS Annual Meeting*, Vol. 34.
- Parsons, S. (1994) Some qualitative approaches to applying the Dempster-Shafer theory *Information and Decision Technologies*, 19, 321-37. (URL: <http://www.csc.liv.ac.uk/~sp/publications/journals/idt.html>)
- Petersen, T. (1994) *Guide to Indexing With the Art and Architecture Thesaurus* 2nd ed. Oxford: Oxford University press.
- Ponceleon, D., Srinivasan, S., et al. (1998) Key to effective video retrieval: effective cataloging and browsing. (Proceedings ACM Multimedia 1998, 99-107) (URL: http://www.acm.org/sigs/sigmm/MM98/electronic_proceedings/ponceleon.html)
- Quintana, Y. (1997) Organization and retrieval in a pictorial digital library, 13-20 in: *Proceedings of the 2nd ACM International Conference On Digital Libraries, July 23-26, Philadelphia, PA*, edited by R.B. Allen and E.M. Rasmussen. New York: ACM.

- Rasmussen, E.M. (1997) Indexing images, 169-96 in: *Annual Review of Information Science and Technology*, Vol. 32, edited by M.E. Williams. Medford: ASIS/Information Today, Inc.
- Röper, M., Hermes, T., et al. (1998?) Video retrieval with the ImageMiner system, 14p. (Postscript from: <http://www.tzi.de/grp/ag-ki/papers/papers.html>)
- Rowley, J.E. and Farrow, J. (2000) *Organizing Knowledge: An Introduction to Managing Access to Information* 3rd ed. Aldershot: Gower.
- Rui, Y., Huang, T.S., et al. (1997) Image retrieval: past, present and future, (PDF from: <http://citeseer.nj.nec.com/192987.html>)
- Rui, Y., Huang, T.S., et al. (1999) Image retrieval: current techniques, promising directions and open issues *Journal of Visual Communication and Image Representation*, 10, 39-62. (PDF from: <http://www.idealibrary.com, article id jvci.1999.0413/>)
- Sarasan, L. (1984) Visual content access: an approach to the automated retrieval of visual information, in: *Automatic Processing of Art History Data and Documents*, Vol. 2, edited by L. Corti.
- Sato, T., Kanade, T., et al. (1998) Video OCR: indexing digital news libraries by recognition of superimposed caption *ACM Multimedia Systems*, (February) (PDF from: http://www.ri.cmu.edu/pubs/pub_2723.html)
- Shatford, S. (1986) Analyzing the subject of a picture *Cataloguing and Classification Quarterly*, 6(3), 39-62.
- Singhal, A. and Pereira, F. (1999) Document expansion for speech retrieval, 34-41 in: *Proceedings of ACM SIGIR, August 1999, Berkeley*. (PDF from: <http://citeseer.nj.nec.com/pdf/133381>)
- Smeaton, A.F. and Quigley, I. (1996) Experiments on using semantic distances between words in image caption retrieval, 174-80 in: *SIGIR '96: Proceedings of the Association for Computing Machinery Special Interest Group On Information Retrieval; 1996 August 18-22, Zurich*, edited by H.-P. Frei, D. Harman, et al. New York: ACM.
- Smith, J.R. (1999) Digital video libraries and the Internet *IEEE Communications Magazine*, 37(1), 18p. (URL: <http://www.comsoc.org/pubs/free/private/1999/jan/Smith.html>)
- Srihari, R.K. (1997) Using speech input for image interpretation, annotation, and retrieval, 140-56 in: *Digital Image Access and Retrieval: Papers Presented At the 1996 Clinic On Library Applications of Data Processing, March 24-26, Urbana-Champaign, IL*, edited by P.B. Heidorn and B. Sandore. Urbana-Champaign: University of Illinois.
- Srihari, R.K. and Zhang, Z. (1999) Exploiting multimodal context in image retrieval *Library Trends*, 48(2), 496-520.
- Sutcliffe, A., Hare, M., et al. (1997) Empirical studies in multimedia information retrieval, in: *Intelligent Multimedia Information Retrieval*, edited by M.T. Maybury. Menlo Park: AAAI/MIT Press.

- Svenonius, E. (1994) Access to nonbook materials: the limits of subject indexing for visual and aural languages *Journal of the American Society for Information Science*, 45(8), 600-606.
- Travis, I. (1998) From "storage and retrieval systems" to "search engines": text retrieval in evolution *ASIS Bulletin*, (April/May), 2p. (URL: <http://www.asis.org/Bulletin/Apr-98/travis.html>)
- Turner, J.M. (1995) Comparing user-assigned terms with indexer-assigned terms for storage and retrieval of moving images, 9-12 in: *ASIS '95: Proceedings of the ASIS 58th Annual Meeting, 1995 October 9-12, Chicago*, Vol. 32, edited by T. Kinney. Medford: Information Today, Inc.
- van Rijsbergen, C.J. (1999) Retrieval through explanation: an abductive inference approach to relevance feedback. (LIC Grant RE/093) (URL: <http://www.dcs.gla.ac.uk/ir/projects/explanation/>)
- Vieira, M.T.P., Biajiz, M., et al. (1999) Metadata for content-based search on an MHEG-5 Multimedia Objects Server, in: *Proceedings of the Third IEEE Meta-Data Conference, Bethesda, April 1999*. (URL: <http://www.computer.org/proceedings/meta/1999/papers/MVieira.html>)
- Voorhees, E. and Garofolo, J. (2000) The TREC Spoken Document Retrieval track *Bulletin of the American Society for Information Science*, 26(5), 3p. (URL: <http://www.asis.org/Bulletin/June-00/voorheesgarofolo.html>)
- Wactlar, H.G., Kanade, T., et al. (1996) Intelligent access to digital video: Informedia Project *IEEE Computer*, 29(5), 46-52. (URL: <http://www.computer.org/computer/dli/r50046/r50046.htm>)
- Witbrock, M.J. and Hauptmann, A.G. (1997) Speech recognition and information retrieval: experiments in retrieving spoken documents, 9p. (URL: <http://www.cs.cmu.edu/afs/cs/user/alex/docs/idvl/slt97.html>)
- Witbrock, M.J. and Hauptmann, A.G. (1998) Speech recognition for a digital video library *Journal of the American Society for Information Science*, 49(7), 619-32.
- Wold, E., Blum, T., et al. (1999) Classification, search and retrieval of audio, 12p in: *CRC Handbook of Multimedia Computing*. (URL: <http://www.musciefish.com/crc/>)
- World Wide Web Consortium (2000) Synchronized Multimedia Integration Language (SMIL 2.0) Specification. (URL: <http://www.w3.org/TR/smil20/>)
- Wu, C.-H., Miller, R.J., et al. (1998) Querying multimedia presentations *Computer Communications*, 21, 1212-25.
- Yoshitaka, A. and Ichikawa, T. (1999) A survey on content-based retrieval for multimedia databases *IEEE Transactions of Knowledge and Data Engineering*, 11(1), 81-93.
- Young, S., Sparck-Jones, K., et al. (1997) Video mail retrieval using voice. (URL: <http://www-svr.eng.cam.ac.uk/Research/Projects/vmr/vmr.html>)

Appendix: Synchronized Multimedia Integration Language (SMIL) A WWW Bibliography

NB This bibliography is available online at
<http://www.mmu.ac.uk/h-ss/cerlim/projects/synchro/smil-bib.htm>

A. HyperText Markup Language @ W3C

W3C - World Wide Web Consortium

Home Page

<http://www.w3.org>

HyperText Markup Language

Home Page

<http://www.w3.org/MarkUp/>

HyperText Markup Language Activity Statement

Description of the current state of play regarding developments in the HyperText Markup Language (HTML), covering SMIL, HTML, XHTML, XML, Stylesheets (CSS), SVG Graphics and MathML.

<http://www.w3.org/MarkUp/Activity.html>

HTML 4.01 Specification

W3C Recommendation 24 December 1999

<http://www.w3.org/TR/1999/REC-html401-19991224/>

XHTML™ 1.0: The Extensible HyperText Markup Language

"A Reformulation of HTML 4 in XML 1.0 W3C Recommendation 26 January 2000"

<http://www.w3.org/TR/xhtml1/>

B. SMIL @ W3C

W3 Synchronized Multimedia

Home Page, includes developments, tools, players, links and history.

<http://www.w3.org/AudioVideo/>

Synchronized Multimedia Integration Language (SMIL) Boston Specification

Current Version: W3C Working Draft 22 June 2000

<http://www.w3.org/TR/smil-boston/>

Accessibility Features of SMIL

"This document summarizes the accessibility features of the Synchronized Multimedia Language (SMIL), version 1.0"

<http://www.w3.org/TR/1999/NOTE-SMIL-access-19990920/>

Synchronized Multimedia Integration Language Document Object Model

"This specification defines the Document Object Model (DOM) specification for synchronized multimedia functionality."

<http://www.w3.org/TR/smil-boston-dom/>

W3C Issues First Public Draft of Synchronized Multimedia Integration Language (SMIL)

Press release: 6 November, 1997 -- The World Wide Web Consortium (W3C) today announced the first public working draft of Synchronized Multimedia Integration Language (SMIL; pronounced "smile").

<http://www.w3.org/Press/SMIL>

The World Wide Web Consortium Issues SMIL 1.0 as a W3C Recommendation

Press release: 15 June, 1998 -- "Leading the Web to its full potential, the World Wide Web Consortium (W3C) today released the Synchronized Multimedia Integration Language (SMIL; pronounced "smile") specification as a W3C Recommendation, representing cross-industry agreement on a wide range of features for putting multimedia presentations on the Web."

<http://www.w3.org/Press/1998/SMIL-REC>

Also: **Testimonials** at <http://www.w3.org/Press/1998/SMIL-REC-test>

World Wide Web Consortium Issues First Working Draft of SMIL Boston

Press release: 3 August 1999 -- "Leading the Web to its full potential, the World Wide Web Consortium (W3C) today releases the first public working draft of Synchronized Multimedia Integration Language (SMIL, pronounced "smile"), known as SMIL Boston."

<http://lists.w3.org/Archives/Public/www-smil/1999JulSep/0037.html>

C. Microsoft

Closed Captions for Web Multimedia

Microsoft® Synchronized Accessible Media Interchange (SAMI) simplifies captioning for developers, educators, and multimedia producers and designers who will now find it easier to make their work more universally accessible.

<http://www.microsoft.com/enable/sami/default.htm>

Spice Up Your Web Pages with HTML+TIME by Debbie Newman, Microsoft Corporation May 2000

This article provides an introduction to HTML+TIME in Microsoft Internet Explorer 5.5.

<http://msdn.microsoft.com/workshop/author/behaviors/htmltime.asp>

Internet Explorer 5.5 with SMIL-Boston support

From: Pablo Fernicola

To: "www-smil@w3.org"

Date: Fri, 17 Dec 1999 11:36:00 -0800

Subject: Beta of Internet Explorer 5.5 with SMIL-Boston support is available

<http://lists.w3.org/Archives/Public/www-smil/1999OctDec/0013.html>

D. Demos/Tutorials

The SMIL Tutorial

A tutorial to demonstrate SMIL capabilities using the Java-based SMIL Player named SOJA by [HELIO](#). Includes introduction to SMIL and step by step guide.

<http://www.helio.org/products/smil/tutorial/chapter1/index.html>

The WebDeveloper.com Tutorial: RealSystem G2 & SMIL By Scott Clark

"In this tutorial, we're going to focus on creating presentations with SMIL and RealPix."

http://www.webdeveloper.com/advhtml/advhtml_tutorial_G2SMIL.html

The WebDeveloper.com Tutorial: Realtext and SMIL By Scott Clark

"In this tutorial, we'll cover RealText, and show you how to use it along with RealPix in your SMIL presentations."

http://www.webdeveloper.com/advhtml/advhtml_tutorial_G2_RealText.html

Learn SMIL with a SMIL

A tutorial plus examples and comparisons by 1999 Jose Ramirez.

<http://www.empirenet.com/~joseram/>

WebDeveloper.com ® Tutorial: SMIL Tools To Get the Job Done By Scott Clark

"Here are the tools you need to put a SMIL on your visitor's faces. By now you've got a pretty good understanding of SMIL, what it can do, and how it is created."

http://www.webdeveloper.com/advhtml/advhtml_tutorial_SMIL_tools.html

Webreview.com - RealSlideshow Plus Simplifies Streaming Media

"RealSlideshow Plus is SMIL-compliant—sort of. The "engine" file seems up to spec, but the audio and image files are both proprietary. In theory you could play a Slideshow Plus on a generic SMIL player, but you wouldn't be able to see any images or hear any audio. Is this an issue? Hard to tell, considering that the RealAudio G2 Player is everywhere."

<http://webreview.com/pub/1999/10/08/feature/index2.html>

Developer.com: Tech Workshop: Working with SMIL

A short tutorial by John Maxwell Hobbs. "Just how easy is easy? We'll begin with a simple slide show of my trip to the Great Wall of China (example 1). This was done in just 16 lines."

http://developer.earthweb.com/journal/techworkshop/092498_smil1.html

How to Create a Simple Digital Story by Derrick Story

"Even the simplest digital story can move and persuade if done properly. Organization and execution are the keys to success. Chances are you have many of the tools right now that you need to create a simple, but effective digital story."

<http://webreview.com/pub/1999/09/24/feature/index3.html> and [Part Two](#)

SMIL at work by Jeff Rule, March 1999

The SMIL Primer, "This article provides an overview of SMIL."

<http://webreview.com/wr/pub/1999/03/12/feature/index.html>

The RealSystem Production Guide

"RealSystem gives you the power to stream compelling multimedia presentations over a network. It includes RealServer, the most advanced streaming media server available, along with RealPlayer and RealPlayer Plus, the world's most popular desktop applications for playing streaming media clips. This manual will help you produce your multimedia presentation, whether a simple video on your home page or a multimedia extravaganza."

<http://service.real.com/help/library/guides/production8/realpgd.htm>

This guide includes:

Chapter 6: Assembling a Presentation with SMIL

<http://service.real.com/help/library/guides/production8/htmlfiles/smil.htm>

SMIL Quick Reference

<http://service.real.com/help/library/guides/production8/htmlfiles/smilref.htm>

Chapter 7: Extending SMIL

<http://service.real.com/help/library/guides/production8/htmlfiles/smilext.htm>

E. Other Multimedia Projects

Projet Opéra

"This project is interested in electronic documents, such as technical documentation, hypertext, and multimedia. It studies document models that take into account logical structures, graphical presentation and multimedia contents. It also develops editing techniques based on these models. The long-term goal is to design and build an editing environment for developing and maintaining large, complex multimedia documentation."

<http://www.inrialpes.fr/opera/>

Captioning and Audio Description on the Web

NCAM (CPB/WGBH National Center for Accessible Media) has developed methods to provide captioning and audio descriptions for Web-based multimedia through the use of QuickTime, SMIL and SAMI.

<http://www.wgbh.org/wgbh/pages/ncam/webaccess/captionedmovies.html>

F. SMIL Web Sites and Lists of SMIL Resources

JustSMIL

Part of the Streaming Media World site, JustSMIL is probably one of the most comprehensive sites currently offering a good selection of tutorials, reviews, links and a gallery.

<http://smw.internet.com/smil/>

Haznet's Fallout Shelter

A 'web designers' site, with a small list of SMIL links under the Fall Out section. Each link is rated with the excellent Geiger Meter.

<http://www.hudziak.com/haznet/>

The CWI SMIL Page

A page of SMIL links produced by CWI which is the National Research Institute for Mathematics and Computer Science in the Netherlands.

Site in English.

<http://www.cwi.nl/SMIL/>

goSMIL

A Yahoo type directory of Streaming Media related resources provided by [PlayStream](#).

<http://www.gosmil.com/>

SMIL

A site aiming to provide "information on a variety of aspects related to SMIL" by the Computational Mathematics Laboratory of the Department

of Computer Science at Concordia University in Montreal.

<http://indy.cs.concordia.ca/smil/>

WebDeveloper.com

WebDeveloper.com's SMIL Links and Resources By Scott Clark.

http://www.webdeveloper.com/advhtml/advhtml_smil_links.html

XML.com

XML's Resource Guide section for SMIL. Includes "SMIL specifications, white papers, tools, and software."

<http://www.xml.com/pub/Guide/SMIL>

Jeff Rule's Dynamic HTML and SMIL Site

Part of Jeff's site with access to some of the articles he has written for WebReview.com and WDVL.com.

<http://www.ruleweb.com/dhtml/smil.html>

Synchronized Multimedia Integration Language (SMIL)

Includes background and links.

<http://www.inrialpes.fr/opera/people/Nabil.Layaida/smil/smil.html>

Yahoo!

Yahoo's directory listing for SMIL: Computers and Internet>Information and Documentation>Data Formats>SMIL

http://uk.dir.yahoo.com/Computers_and_Internet/Information_and_Documentation/Data_Formats/SMIL/

Open Directory

Open Directory's directory listing for SMIL: Computers: Data Formats: Markup Languages: SMIL

http://dmoz.org/Computers/Data_Formats/Markup_Languages/SMIL/

LookSmart

Directory Listing for SMIL: Computing - Computer Science - Programming - Internet & Scripting - SMIL

<http://www.looksmart.com/eus1/eus317831/eus317876/eus53906/eus65717/eus278683/r?l&>

G. SMIL Mail Lists

www-smil@w3.org Mail Archives

The official W3C SMIL mailing list.

<http://lists.w3.org/Archives/Public/www-smil/>

H. SMIL Articles

Moving to the beat by Lloyd Rutledge, 1999

The Web used to stand still. SMIL gives the Web a sense of timing and adaptation.

<http://www.heise.de/ix/artikel/E/1999/10/058/default.shtml>

The moving picture by Jan Ozer, October 1999

The Dark Side vs. the Real World or the Real Outer Limits.

<http://www.emediapro.net/EM1999/picture10.html>

Toward Synchronized Multimedia on the Web by Philipp Hoschka, Spring 1997

<http://www.w3j.com/6/s2.hoschka.html>

Introduction to SMIL by Jeff Rule, December 1998

"The Synchronized Multimedia Integration Language (SMIL) is a recommendation from the World Wide Web Consortium (W3C) that allows for the creation of time-based multimedia delivery over the web. Based on XML, it allows developers to mix many types of media, text, video, graphics, audio and vector based animation together and to synchronize them to a timeline."

<http://wdvl.internet.com/Authoring/Languages/XML/SMIL/Intro/>. Also [other Jeff Rule articles](#) dealing with RealVideo, RealText, RealPix and RealAudio.

Making Sites SMIL by Steve McCannell, May 5 2000

<http://webreview.com/pub/2000/05/05/features/index01.html>

Are You Smiling About SMIL?

Poll Results, May 2000

<http://webreview.com/wr/pub/2000/05/05/poll/results.html>

SMIL's New Strategy: Modularity by David Sims, Derrick Story, October 1999

"We at Web Review have always been interested in SMIL, partly because it's one of those technologies (like Flash and Animated GIFs) that bridges several of the areas we cover: design, programming, and standards. This week, Kim Brown offers a review of some of the features in the latest working draft of the Synchronized Multimedia Integration Language, SMIL Boston."

<http://webreview.com/wr/pub/1999/10/08/feature/index.html>

The Future of SMIL by Kim Brown, October 1999

"The W3C (World Wide Web consortium) sports the following motto: "Release early, release often." It's a sage maxim to follow when changing the development course of a Web standard. The most recent release of SMIL (SMIL Boston) by the W3C Working Group demonstrates why getting a preliminary model out early is so important." <http://webreview.com/pub/1999/10/08/feature/index3.html>

SMIL - a standard for multimedia (at last!) by Mahesh Shantaram, September 1999

"A very simple language to integrate text, graphics, audio, and video into an online, interactive presentation."

<http://www.ciol.com/content/technology/techbytes/99091402.asp>

Captioning and Audio Description on the Web

"NCAM has been experimenting with ways to provide captioning and audio descriptions for Web-based multimedia through the use of QuickTime, SMIL and SAMI."

<http://www.wgbh.org/wgbh/pages/ncam/webaccess/captionedmovies.html>

Group Makes Strides in Improving Web Access for the Disabled by Brian Hannon, September 1998

XML and Synchronized Multimedia Integration Language are being used to help improve the Web for people with disabilities.

<http://www.wgbh.org/wgbh/pages/ncam/bp/news/webnews4.html>

Making CD-ROM's Multimedia for all users by Tom Wlodkowski, June 1999

"Advances in computer technology have brought most people as close as a mouse click to a wealth of information. It is no longer necessary to flip through volumes upon volumes of an encyclopedia to access a map of Africa, or to search for information about Babe Ruth."

<http://www.wgbh.org/wgbh/pages/ncam/bp/news/cdromnews1.html>

An Interview with the W3C's SMIL Guy, Phillip Hoschkaby by D.C. Denison January 9, 1998

"On Nov. 6 1997, the W3C released the first public draft of the Synchronized Multimedia Integration Language (SMIL, pronounced "smile"), a specification that promises to make it a lot easier for developers to create multimedia content for the Web. Philipp Hoschka is the chair of the W3C Synchronized Multimedia Working Group and the editor of the draft."

<http://webreview.com/98/01/09/feature/interview.html>

**Toward Synchronized Multimedia on the Web by Philipp Hoschka.
World Wide Web Journal, Spring 1997**

"Web technology is limited today when it comes to creating continuous multimedia presentations. For these applications, content authors need to express things like "five minutes into the presentation, show image X and keep it on the screen for ten seconds." More generally, there must be a way to describe the synchronization between the different media (audio, video, text, and images) that make up a continuous multimedia presentation."

<http://www.w3j.com/6/s2.hoschka.html>

Synchronizing the Web: Choreographing multimedia makes developers SMIL by Kim Brown January 9, 1997

"In the movie Broadcast News there's a scene where Joan Cusack races through a television station carrying a crucial piece of videotape that needs to be cued up and played at the exact moment newscaster William Hurt begins reporting about the footage. Vaulting over filing cabinet drawers, blowing past coworkers, she sprints for the video player and....TOUCHDOWN! Television viewers tuned into the broadcast get what they expect to see: A seamless, synchronized presentation of audio and video."

<http://webreview.com/98/01/09/feature/smil.html>

STANDARDS: Time-Based Multimedia Technology Nears Approved Status By Nate Zelnick, Internet World April 13, 1998

"The World Wide Web Consortium (W3C) announced last week that it has moved a technology for building multimedia applications into the final stage before becoming a standard, making it likely that the Synchronized Multimedia Integration Language (SMIL, pronounced "smile") will be a recommendation before summer."

<http://www.internetworld.com/print/1998/04/13/news/19980413-multimedia.html>

Intraware: Tools: Research: Library: Intraware: Position Paper: SMIL Observed by Esteban Kolsky

"Esteban Kolsky's take on SMIL: what it is, why it is, and where it's going -- not to mention who is going to support it and who says they'll support it but have developed their own standard for multimedia that they will push... (hint -- does this sound like ActiveX revisited?). Also included are examples of SMIL code and links to related information."

<http://www.intraware.com/ms/itwr/pospr/SMIL.html>

Two Web Video Captioning Technologies From About.com's Deafness/Hard of Hearing section by Jamie Berke.

"Companies, individuals, and organizations that put video on the web have absolutely no excuse for not making their videos accessible to the deaf and hard of hearing. Both Microsoft and Real Networks have developed captioning technologies for use with digital video. "

<http://deafness.about.com/health/deafness/library/weekly/aa083198.htm>

I. SMIL Software

RealNetworks

Home of the Real Player streaming media player , SMIL compliant.

<http://www.realnetworks.com/>

Real Slideshow Plus - SMIL Authoring tool

RealNetworks says; "Simply drag and drop your digital images onto the interface, crop and crunch images, add voice and text narration for each image, add some background music and press "Generate." Then press "Play" to view your creation."

http://www.realnetworks.com/products/slideshowplus/info.html?src=prdctmn_072600b

Also for RealNetworks:

DevZone - The informational resource for web developers and designers

<http://www.realnetworks.com/devzone/index.html>

SMIL Presentation Wizard User's Guide

<http://service.real.com/help/library/guides/smilwiz/smilwiz.htm>

QuickTime

Apple's multimedia player and developers tool.

<http://www.apple.com/quicktime/>

Developers information.

<http://developer.apple.com/quicktime/>

Information on Quicktime and SMIL

<http://www.apple.com/quicktime/authoring/qtsmil.html>

SMIL at HELIO

HELIO is a French non-profit organization. SOJA is their Java-Based SMIL Player. SOJA stands for SMIL Output in Java Applets. It is able to render SMIL presentations into web pages using HTTP and simple medias.

<http://www.helio.org/products/smil/>

GRiNS Authoring Software

The GRiNS family of multimedia presentation authoring software offers a full range of tools to build compelling presentations using SMIL, the technology for multimedia on the Internet.

You can use GRiNS to create streaming multimedia presentations containing audio, video, text and images for the over 70,000,000 RealSystem G2 players. Or, you can use GRiNS to make "pure" SMIL presentations containing a wide range of media from interactive HTML to complex animations.

<http://www.oratrix.com/GRiNS/>

Fluition by Confluent Technologies

"Fluition is a software tool for the layout and sequencing of streaming multimedia presentations. Fluition was designed to be easy to use but offer all of the power and flexibility that SMIL offers. Fluition eliminates the need for manually writing code and allows full multimedia layout and sequencing capabilities in a visual environment. Fluition can make use of a wide range of media file formats making it possible to create very sophisticated productions limited only by the producer's

imagination."

<http://smilsoftware.com/>

SMIL Composer from Sausage Software

"The new SMIL Composer SuperTool allows you to easily create synchronised multimedia content for RealSystem G2. With its easy point and click WYSIWYG layout interface no knowledge of SMIL code (an implementation of XML) is required. You can easily add available media types, arrange their layout and sequence how they are played in your composition. Then it is a one button press to view your SMIL Code or preview your composition in your RealSystem G2 Player."

<http://www.sausage.com/supertoolz/toolz/stsmil.html>

Media Access Generator (MAGpie)

"Developers of Web- and CD-ROM-based multimedia need an authoring tool for making their materials accessible to persons with disabilities. The CPB/WGBH National Center for Accessible Media (NCAM) has developed such a tool, the Media Access Generator (MAGpie). Using MAGpie, authors can add captions to three multimedia formats: Apple's QuickTime, the World Wide Web Consortium's Synchronized Multimedia Integration Language (SMIL) and Microsoft's Synchronized Accessible Media Interchange (SAMI) format. MAGpie can also integrate audio descriptions into SMIL presentations."

<http://www.wgbh.org/wgbh/pages/ncam/webaccess/magindex.html>

Dreamweaver Extensions by AHEAD

W3C SMIL 1.0 DTD - this browser profile can be downloaded for Macromedia's Dreamweaver to check your SMIL against the standard.

<http://www.thought.co.uk/extensions/browserprofiles.html>

Video 123 from WebKapture, Inc.

Streaming Video editor, can output SMIL format

<http://www.webkapture.com/video123/home.php3>

ExtendMedia's T.A.G. Composer 2.0

A streaming media tool, SMIL output.

<http://tagsoftware.com/>

Experimental SMIL syntax validator

Check your SMIL code.

<http://www.cwi.nl/~media/symm/validator/>

J. Examples

Info and Comms @ MMU

A guided tour of the Department of Information and Communications at Manchester Metropolitan University. Created as an SMIL example for the Synchronisation project.

View the Demonstration at <http://www.mmu.ac.uk/h-ss/cerlim/projects/syncro/demo.html> - Needs Realplayer
<http://www.real.com>

From Real, three examples on the Realsideshow Plus page, plus their own demonstrations.

http://www.realnetworks.com/products/slideshowplus/info.html?src=prdctmn_072600b

From Electric Ladyland, using the Fluition authoring tool.

<http://www.electricleisureland.com/cgi-bin/elland.cgi>