University of Southampton

The Multimedia Thesaurus: An Aid for Multimedia Information Retrieval and Navigation

by

Robert Tansley

A mini-thesis submitted
for transfer of registration from
MPhil to Ph.D

in the
Faculty of Engineering and Applied Science
Department of Electronics and Computer Science

December 10, 1998

Contents

List of Figures

Abstract

In many areas of work, large amounts of information accumulate. Ways to store and access this information have been the subject of much research for the last century and even earlier. The advent of computers has greatly extended the possibilites for information retrieval. However the situation is still not an ideal one; many problems still exist. Many of the problems associated with current multimedia and hypermedia information systems can be addressed by adapting the well-established thesaurus technique.

By using media matching and processing technology, textual terms can be supplemented with representations of other media types. This thesis describes such a multimedia thesaurus and the ways in which it can be used. Traditional thesaurus information retrieval techniques are explored along with existing systems research with similar aims. The textual thesaurus paradigm is extended to take advantage of the wealth of knowledge representation research, allowing artificial intelligence and software agents to aid the searcher. A prototype system implementing a multimedia thesaurus, *MAVIS 2*, is introduced, and methods of evaluation discussed.

1 Introduction

In many areas of work, large amounts of information accumulate. Areas such as history, medicine, research and industry all have need of large information stores such as historical archives or medical databases. Ways to store and access this information have been the subject of much research for the last century and even earlier. The advent of computers has greatly extended the possibilites for information retrieval. Now it is possible to store large amounts of information, and retrieve information far more quickly and independent of location than before.

However the situation is still not an ideal one; many problems still exist. One of the principle problems is finding ways of explaining to a computer exactly what it is you wish to know in such a way that the computer will retrieve all the relevant information you want, but not retrieve any irrelevant information. Such a "perfect" system is still a long way off; however it is possible to go some way to remedy the situation.

Retrieval of textual information is relatively well-established. The medium of text can be easily stored and processed by a computer since it consists of a stream of characters from a limited set of possible characters. It is therefore easy for a computer to compare and manipulate textual information.

Retrieval of non-textual media, however, is a relatively young area and presents a number of more complex problems. Multimedia data is not so trivially representable. For example, an image must be quantised; that is, divided into a matrix of numbers, where each number represents the colour at that location. So when a picture of a real-world situation is stored in the computer, some of the information is lost. A computer also lacks the *a priori* knowledge about the world that humans possess; they cannot "understand" or perceive the image in the same way. A variety of mathematical and other techniques must be used. Even then, two media objects representing the same real world object will not be exactly the same, even in the difference is unobvious to the human eye or ear.

MAVIS 1, a previous research project addresses this problem by using such mathematical techniques to evaluate how similar two media objects are, and using this information to allow the user to retrieve multimedia objects from an information set, and to navigate around this information by following links. The system does not however constitute a complete solution.

Another problem is the fact that different views of an object may exist; for example it is easy for a person to recognise a car from any viewing angle, even though the view of the car is very different from each angle. This problem is however similar in nature to the terminology problem in text; two authors may use different terminology to describe the same concept. The research presented in this thesis draws on this idea.

The use of text thesauri in information retrieval is a firmly established technique. By using media matching and processing technology, textual terms can be supplemented with representations of other media types. This thesis describes a multimedia thesaurus and the ways in which it could be used.

A prototype system implementing a multimedia thesaurus, *MAVIS 2*, is also introduced. *MAVIS 2* is the successor to *MAVIS 1*.

1.1 Overview

Chapter 2 describes possible structures for a multimedia thesaurus, and describes a number of ways in which it can be used to aid the information retrieval and navigation processes. It also covers how such a thesaurus could be initially constructed.

Chapter 3 describes work on the MAVIS 2 project, including an architecture for the data. The chapter also explains how navigation and retrieval operations will be performed. A graphical user interface has been prototyped and is also described.

Lastly Chapter 4 summarises this research and outlines future work, including a time plan and a brief plan of how the techniques described earlier can be evaluated.

Appendix A of this thesis describes existing information retrieval and navigation techniques, and reviews related research publications.

2 The Multimedia Thesaurus

2.1 Introduction

This chapter describes what a Multimedia Thesaurus or MMT consists of and the possible ways it can be used to aid an author or searcher of a multimedia information system. This chapter represents my own work.

2.2 Multimedia Data as Terms

Textual terms used in textual thesauri are symbols representing a particular subject (or *concept*). Saussure termed these symbols “signifiers” [49]. Other media items such as images or video contain components that represent its meaning in a similar manner. These components can be considered as “terms” in a thesaurus, each of which pertain to one or more concepts. In the remainder of this chapter these multimedia terms are called *representations*.

These representations are not the only part of the thesaurus. Consider a controlled language thesaurus stripped of all but the preferred forms of a term; each concept is represented only once, as are semantic relationships between these concepts. The stripped terms are used only as a means of reaching this network of concepts from a query and are called “lead-ins” in traditional thesaurus terminology. This terminology has been adapted for the MMT. Multimedia understanding and matching technology in the classic content-based retrieval sense can be used to match those lead-ins that are non-text. Provided such technology exists for a particular type of media, the MMT can integrate that media type.

It is likely that only some properties of the media are representative of the concept to which it is associated. For example, a number of images of cars may lead in to the concept “car”. Cars come in many colours — the colour of the car is not an important factor whereas the shape of the car is. In this case a lead-in consisting of an image of a car is *qualified* — one or more properties of the media (in this case, the shape) is indicated as being relevant to the concept "car". This is analogous to the text "red car" appearing in a document. Only the word "car" is considered as part of the lead-in; the fact that the car is red is irrelevant to the concept. In addition it may be appropriate to specify an application domain-specific matching algorithm that should be used.

The MMT described consists of two parts. Firstly there is the network of concepts, connected by traditional thesaurus relationships. Secondly there is the set of multimedia lead-ins that queries are matched against. Since the matching of non-text media is inexact and often domain restricted, the strength of the match can be considered indicative of the likelihood that a particular query is relevant to the concept to which the lead-in is associated.

While it may be natural to think of the concepts in the concept network as the preferred form of the concept, these preferred forms may still be used for matching during indexing and retrieval. It is therefore convenient to think of the preferred representation of a concept as another lead-in, and the concepts as abstract entities. The preferred representation must be tagged as such, since when navigating the concept network, each concept must be represented on-screen. It should always be a textual term since that is easily displayable, and often the most concise and unambiguous way of representating a concept. Visualising a network using many different media types may also prove complex and confusing.

2.3 Identifying Concepts Relevant to a Term

One of the basic operations required in a thesaurus system is the identification of where a query term appears in the thesaurus. In the case of text this is trivial. In the general case, media-specific matching techniques (see section A.4.1) must be used to find the best matching representation(s) of concepts in the MMT.

Given a query object, the matching technology is used to match the query with all the representations in the MMT. Due to the uncertain nature of multimedia matching, each representation in the MMT of the same media type as the query will have a match strength.

To determine which concept the query represents can be achieved a number ways:

- The concept to which the best-matching representation leads in can be automatically selected. This is however fairly unreliable, unless the strength of the best match is both a strong one and significantly higher than other matches. This will work well with text.

- If a query matched well with more than one representation of a concept, then the likelihood of an association is increased. Such a method assumes that each concept has a comparable (high) number of associated representations.

- The system could present a ranked thresholded list of the best matching representations, together with the concepts they represent. The user can then see which representations matched well, and choose a concept they feel is most appropriate. This allows the system to make use of the user's knowledge.

- Some intelligent agent process could return results. This could be combined with the above methods. Detailed study of these is outside the scope of this research.

Once the most appropriate concept for a query is determined, the query selection may added to the thesaurus as an additional representation of that concept. Some user interaction may be required to determine what it is about the query object that is representative of the concept.

The process can also be used to assist the user when initially entering objects into the system. The object can be searched for representations held in the MMT, and a list of concepts likely to be relevant to the object presented. This fulfills key objectives 4 described in section A.1.1.

2.4 Uses of the Multimedia Thesaurus

There are a number of ways an MMT could be structured and used to aid the multimedia information retrieval and navigation process. These methods are explained below.

2.4.1 Lists of Synonyms

A simple MMT consisting of just lists of equivalent representations (synonyms) would be simple to construct and useful in two ways.

2.4.1.1 Query Expansion

Once the concept most relevant to a query is found, other representations of that concept can be used to augment the query, regardless of their media type. In this way it is possible to retrieve documents of media types other than that of the query.

The user may start a query using a selection from an image of an office chair for example. If a strong match is found with an image leading in to the concept "office chair", a text representation of that concept can be used to start CBR operations retrieving documents containing text. If content-based hypermedia linking services are available, each representation can be used to find appropriate links. In the above example the audio representation of the concept could be used to offer links whose audio source anchors match that representation.

This facility fulfills key objectives 1 and 3 described in section A.1.1.

2.4.1.2 Link Source Anchor Normalisation

This method involves "switching" a source anchor when it is authored. If the source anchor of a link is (or matches well with) a representation in the MMT, the source anchor can be "switched" to the preferred representation of that concept. The source anchor can then be added as a representation of the concept.

During navigation, if the user's selection matches well with a representation of a concept, the preferred representation of that concept can be used to find all links authored on any representations of that concept.

For example, a user might author a link using an image of a car as the source anchor. This matches well with another image of a car in the MMT, a representation of the concept *car*. The source anchor is switched to "car" before the link is stored, and the original source anchor is added to the MMT as another representation of the concept car.

This facility also fulfills key objectives 1 and 3 in section A.1.1.

2.4.2 A Connected Thesaurus

If the concepts are connected with semantic relations, more possibilities for aiding a searcher exist.

2.4.2.1 Fan-out

The query expansion method described above can be supplemented by adding representations of semantically related concepts to the query. For example a query might consist of the textual term "car". Representations of that and narrower concepts can be used to retrieve images of cars; representations of narrower concepts of "car" are also likely to be cars, but specialised instances rather than general.

The concepts and representations to be used can be determined in a number of ways.

- The user should be able to decide which media types are used.
- Some form of semantic distance measure can be used to determine which concepts' representations should augment the query.
- The types of relation traversed can be constrained. A particularly useful form of this would be to augment the query with representations of the initial and narrower concepts.
- A "concept selector" tool would allow the user to manually select the concepts (and representations) used.

This is one way of implementing key feature 2 of the MMT.

2.4.2.2 Relevant Documents

In a traditional thesaurus as used in text information retrieval, each concept (set of equivalent terms) is used to index a document collection. With the preferred form of each term is held a list of documents relevant to that term. This technique can adapted to the MMT very simply. With each concept can be held a list of documents (or rather, media objects) relevant to the concept. Once a concept is reached from a query, the set of these objects can be presented to the user.

To avoid confusion, and to keep in line with traditional thesaurus terminology, these associated media objects (which are *not* representations of the concept) are termed *relevant documents.*

2.4.2.3 Thesaurus Navigation

Starting with the concept originally determined to be relevant to the query, the user can be presented with a textual or graphical visualisation of the thesaurus concept network. The user can navigate around this, reaching concepts by traversing relationships. At each stage the user must be able to see representations of the concept, and also to see what relevant documents are available and view them.

The type of relations should be clear to the user. This enables the user to refine or expand their query by traversing the *broader concept* or *narrower concept* relations for example.

This can be considered as another form of hypermedia navigation. Rather than navigating between documents, the user navigates between the higher level concepts to which the documents pertain.

Thesaurus navigation is another way of implementing key feature 2 of the MMT.

2.4.2.4 Narrowing the Scope of a Query

If an MMT has both a comprehensive set of representations and indexes all necessary relevant documents to a task, it can be effectively used to narrow the scope of a search, reducing the amount of irrelevant information returned as speeding up the search process.

The user is allowed to select some portion of the MMT. Once this portion is identified, only representations and relevant documents associated with the selected concepts are used in search and navigation operations. For example, if the user is only interested in cars, they can specify that only the concept "car" and narrower concepts are considered. The user can then submit a query consisting of just an abstract property such as the colour red, only pictures of red cars will be retrieved. To retrieve red cars using another method, the user would need to include a representation of a car in the query, and even then only similarly-shaped views of cars would be retrieved.

There are a number of ways the user can select the concepts.

- Select concepts manually from a "concept selector."
- Select a concept and all narrower concepts. Another type of relationship could also be used.

- Select a concept and its surroundings (defined by a semantic distance measure such as number of relations traversed.)

2.4.3 Extending Multimedia Thesaurus Functionality

The layer of concepts in the MMT can be supplemented to further enhance its capabilities.

2.4.3.1 A Richer Set of Semantic Relationships

A traditional thesaurus supports only two types of semantic relationship other than equivalence, *broader/narrower* and *related*. The *broader/narrower* relationship is effectively a single, directed relationship between two concepts. It indicates that the scope of one concept is encompassed by the scope of another. The *related* relationship is used to indicate all other kinds of association between concepts, and is a single bidirectional connection between those concepts.

The two types do not give much information about the relationship. This limits the amount of semantic information that can be used to supplement a query. In addition, the ways in which the concepts and relationships can be viewed is limited. For example, in a typical faceted thesaurus, the searcher can specialise and generalise in each facet, and cross facets using *related term* connections. It may be the case that other natural structures exist in the thesaurus, but these cannot be exploited by either the searcher or the retrieval system.

There is no reason why the concept network in the MMT need be so restricted. For each individual application, there is likely to be a restricted set of appropriate relationships, each of which have their own structure. For example, one thesaurus might contain an *is_type_of* hierarchical structure, and another might require arbitrary spatial relationships.

Each MMT must therefore conform to a *relationship schema*. This consists of all the relationship types allowed in the concept network, together with the structural constraints that should be enforced on each of these relationships.

Relationship constraints assist the constructor of the concept network in that they reduce the likelihood of inconsistencies and inaccuracies. However such constraints alone do not guarantee a well-authored conceptual network; the network must be richly

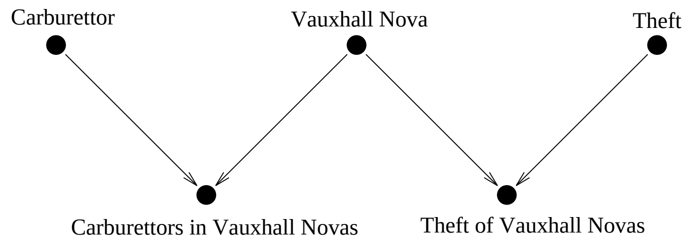

Figure 2.1: Search Space using MMT Concepts

connected, and there must be no isolated subnetworks [51].

2.4.3.2 Weighted Relationships

Connections in the concept layer indicate a relationship between the two concepts. The strength of the relation may vary; for example the concepts "Vauxhall Nova" and "Vauxhall Nova Pearl" are strongly related; the concept "carburettor" and "Vauxhall Nova" are related but not as strongly. The assigning of weights indicating the strength of a relationship will enhance in particular in deciding the extend of 'fan-out" during query expansion.

2.4.3.3 A Search Space

A hyperindex or search space can be constructed using an MMT. Each concept in the MMT can be connected to another in the MMT to provide a more precise way of specifying a searcher's need. For example, the concepts "Vauxhall Nova" and "carburettor" can appear together as "Carburettors in Vauxhall Novas." Documents relevant to both can be indexed by this compound concept. The user can navigate a search space such as that shown in figure 2.1 with fewer irrelevant documents appearing at many nodes than when just navigating MMT concepts.

This method is still under research. It may involve the use of complex text processing techniques. Additionally it might not be easily applicable to the multimedia domain.

2.5 Construction

Constructing a multimedia thesaurus involves up to four stages:

2.5.1 Constructing the Relationship Schema

If the MMT is to use non-traditional thesaurus relationships, the relationship schema as described in the previous section must be designed. The information to be specified for each relationship is:

- A name for the relationship. In the case of relationships where traversal in either direction holds different meaning (for example the *broader/ narrower* relationship) names for both directions should be specified.
- Structural constraints for the relationship. This may take the form of selecting one from a set of preset types (e.g. hierarchy, graph, acyclic graph etc.) or specifying a set of declarative rules. If may also be appropriate to specify whether or not the relationship is a specialisation/ generalisation (as is the case with *broader/ narrower* — in these cases, since a general term encompasses the scope of the specialisations, lead-ins and documents relevant to the specialisations are implicitly also associated with the more general concept.
- Other constraints. For example in a *broader/ narrower* relation hierarchy, it may be inappropriate for a document to be associated with a term when it is also associated with a narrower form of that term (see figure 2.2. Documents are best associated with the narrowest appropriate term [1]. For example a document about a particular type of bird (for example, a magpie) should not be connected to the broad concept "bird", but rather to the appropriate narrower concept.

2.5.2 Constructing the concept network

This may be done by hand or with assistance. Construction by hand is laborious but may produce the best results. Some construction tool that allows the user to create and connect concepts must therefore be provided. This tool must enforce the constraints that are specified in the relationship schema, and must ensure that the resulting network contains no isolated subnetworks. It will also be necessary to allow the thesaurus constructor to specify at least one textual representation for each concept for visualisation purposes.

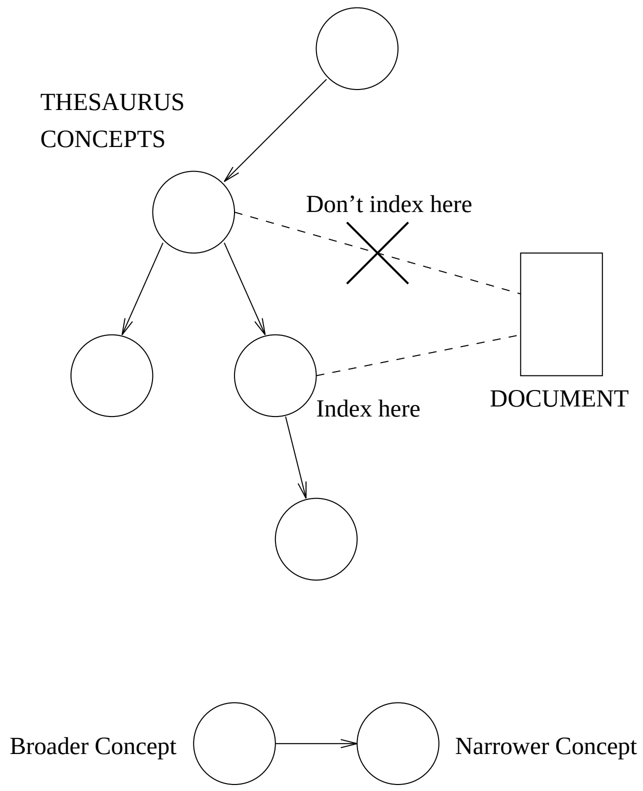

Figure 2.2: Associating documents with narrowest appropriate term

It may be the case that a classification schema for a particular application already exists. In this case the thesaurus constructor will use the same construction tool.

Automated techniques may also be employed with suitable software tools. Existing automated classification techniques can be augmented with media understanding and processing technology, enabling relationships between entities in non-text media to be inferred. Intelligent agents and artificial intelligence techniques may also be employed — their use is an ongoing field of research [35]. It is likely that some user interaction would be required, and the resulting network should be checked by the constructor.

2.5.3 Adding representations for concepts

Each concept is already likely to have a textual representation from the previous stage, and if automated techniques are available the concepts may also have representations of other media types. The thesaurus constructor can at this stage add further representations of concepts. This may be done in a number of ways:

1. Selecting a piece of media from a document, and identifying relevant properties (qualifying) it.

2. Entering an "anonymous selection" — that is, a piece of media that has not come from a document. This may take the form of typing in some text. In the case of other media, this can take the form of identifying a property or properties of that media without actually requiring an example. For example, the constructor may select a colour using RGB sliders.

3. If a thesaurus or some other organised media collection already exists, an automated process may extract other representations of a concept from it.

2.5.4 Indexing the document collection

While this can be done manually with a suitable authoring tool, it is likely that an automated process will provide good results, provided each concept has appropriate representations (that is, every representation of a concept appearing in the document collection is held in the thesaurus.) The document collection can be searched for each representation in the thesaurus, and each document in which that representation appears (or is confidently considered to appear) is connected to the appropriate concept, subject to the constraints laid out in the relationship schema.

The constructor may be involved in a similar semi-automated process, and may want to check and/or alter the resulting thesaurus.

2.6 Concept Browser Prototype

So far the only prototyped component of the MMT is a concept browser. This enables the user to navigate around the concepts in an MMT, and see the representations of and relevant documents associated with each concept. Shots of two of the prototypes in operation are shown in figures 2.3 and 2.4. The first prototype shows concepts (or rather, their preferred representations) as text in lists, the other presents a simple graphical visualisation. Preliminary discussion supposed that the graphical visualisation style is more intuitive.

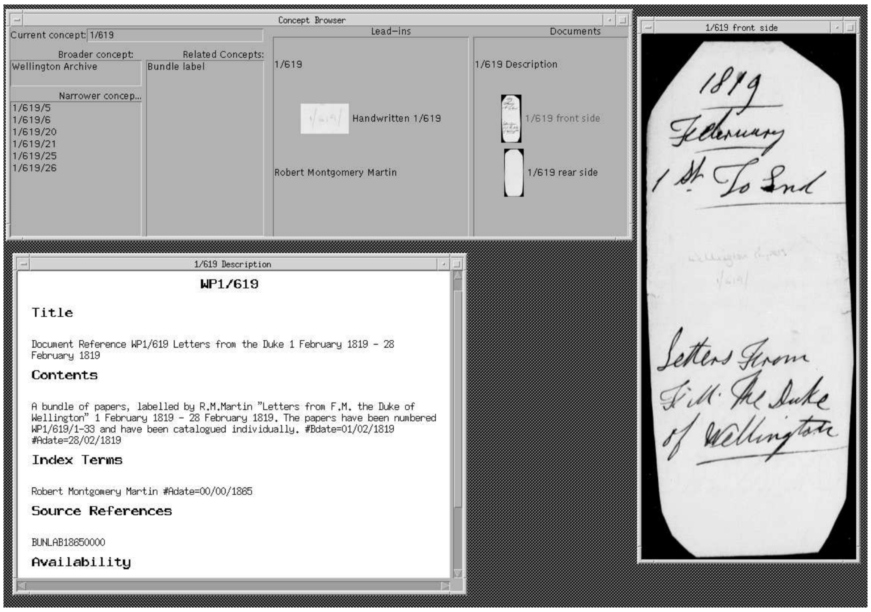

Figure 2.3: Text-based Concept Browser Prototype

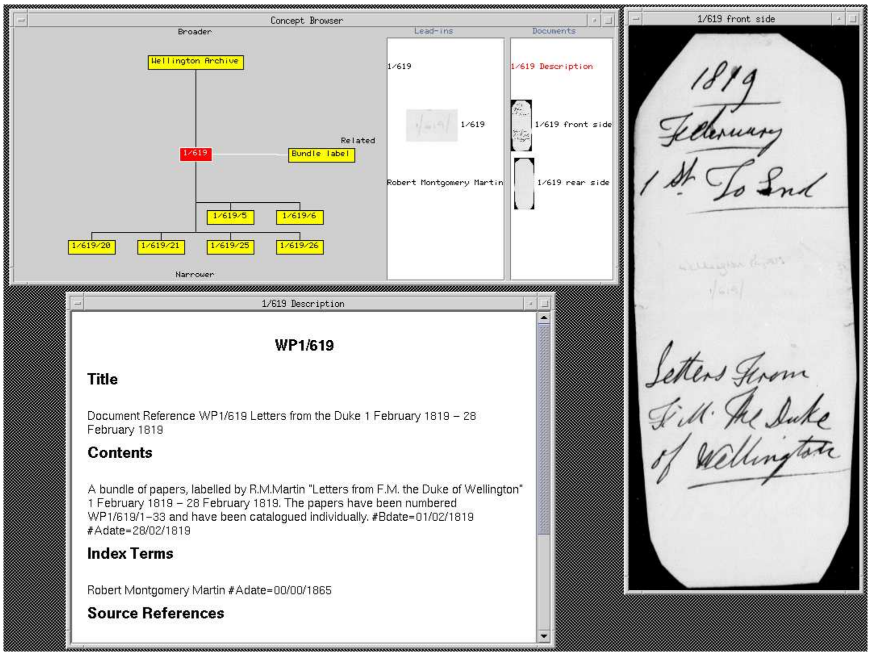

Figure 2.4: Graphical Concept Browser Prototype

3 The MAVIS 2 Open Hypermedia System

3.1 Introduction

This chapter describes MAVIS 2, a prototype open hypermedia system supporting a multimedia thesaurus feature and intelligent agent support. This work has been produced in collaboration with other members of the MAVIS 2 development team, Mark Dobie, Dan Joyce and Paul Lewis.

3.2 Four Layer Data Architecture

Information in MAVIS 2 is divided into four layers. An overview can be seen in figure 3.1. Each layer is described in the following sections.

3.2.1 Raw Media Layer

This consists of all of the raw media usable within the system. Every piece of actual media is held at this layer, though to be usable by the rest of the system it must be referenced by a selection in the selection layer.

3.2.2 Selection Layer

The selection layer of selections from raw media. Each selection will typically consist of a specification of a piece of media from the raw media layer. Any signatures that have been calculated will be held here too. Each selection may contribute to one or more selection expressions.

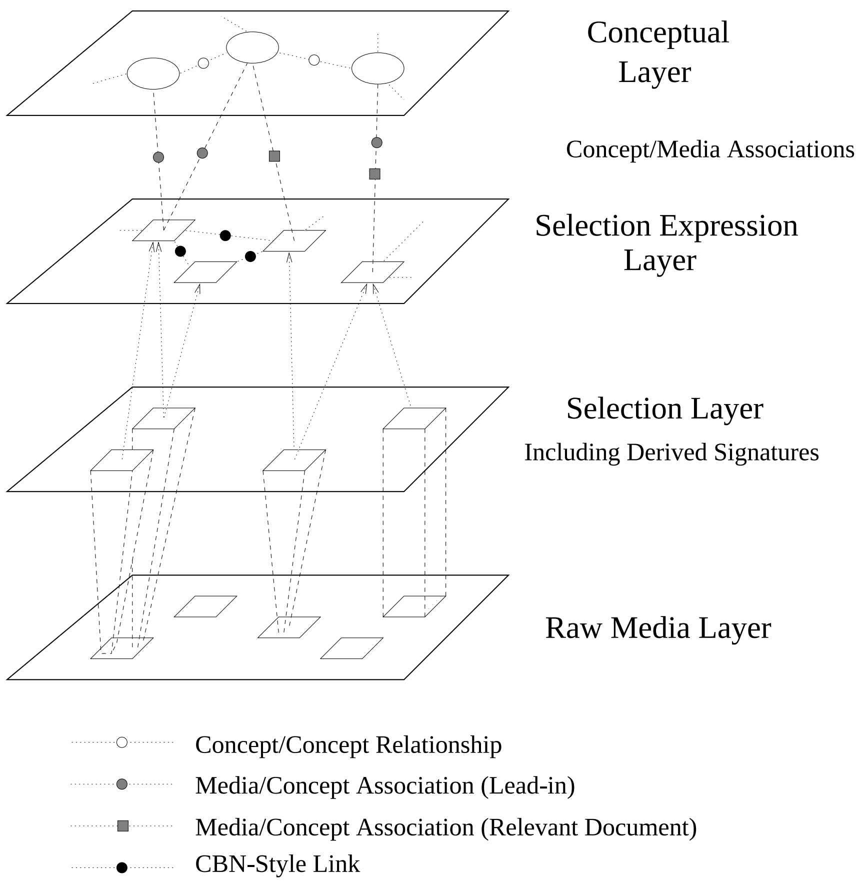

Figure 3.1: Four-Layer MAVIS 2 Data Architecture

3.2.3 Selection Expression Layer

This layer consists of selection expressions and links between them. Selection expressions pertain to one or more selections from the selection layer, and information about which properties of each selection are relevant. Simple combinations of selections may be constructed by means of a simple GUI, for example *the colour of selection S1* and *the shape of selection S2*. Most selection expressions will consist of a single selection with no particular property specified as relevant (in which case *all* properties are considered relevant.) Thus a selection expression can be thought of as a way of specifying a set of signatures or properties identifying an object.

Selection expressions may be source anchors of links. The destinations of these links may be either a concept in the conceptual layer, or another selection expression. The former case is described in the next section. The latter allows traditional multimedia content-based navigation. Typically the destination of such a link will be a selection expression pertaining to a single piece of media. Such a link can have arbitrary key-value attributes associated with it for application-specific purposes.

3.2.4 Media/Concept Associations

These associations connect the conceptual layer with the selection expression layer. Arbitrary key-value pairs can be attached, and the association can be weighted. Although they are essentially bidirectional, the way they will typically be used is determined by their type. The association may be a *lead-in* (a representation of a concept in the MMT) or a *relevant document.*

3.2.5 Conceptual Layer

This consists of abstract representations of concepts. They do not have associated keywords stored at this layer. No media of any kind is held at this layer.

Relationships between these concepts may be traversed in either direction, although the context of the relationship may differ in each direction, for example specialisation in one direction and generalisation in the other. These relationships can have arbitrary key-value pairs associated with them, allowing application-specific information to be stored such as weightings. Although this layer by default will use classic thesaurus-style relationships (*broader/narrower, related)* a richer set of semantic relationships can be expressed subject to a *relationship schema* (see section 2.4.3.1).

Although this is the only layer exclusively used by Multimedia Thesauri, an MMT will require the use of the other four layers as well. However an MMT can be easily disabled by switching out this layer.

3.3 Queries in MAVIS 2

Navigation in MAVIS 2 takes the form of queries. Typically a user will provide a selection expression (already existing or specifically created,) information about what type of result they want (whether they want concepts to be returned, other selection expressions, or links) and how the MMT and/or agents should be involved.

The query is processed and the results presented to the user. The user can then either follow links, view other selection expressions or view and browse concepts in the MMT.

A query consists of three parts, described in the following sections.

3.3.1 A Selection Expression

The user specifies a *query selection expression* that provides the signatures against which all selection expressions (be they lead-ins, source anchors of links or neither) will be matched. The selection expression interface will allow the user to express signatures which, if matched, indicate decreased likelihood that the object matched will be useful to the user. In other words, the user can effectively say "I'm not interested in objects like this."

After the query has been processed this query selection expression will be modifiable, allowing the user to perform relevance feedback iterations. For example if an object matched well with the query selection expression, but the user decides the object is not relevant, the object can be appropriately attached to the query selection expression.

3.3.2 A Results Specification

Usually the user will have some idea of what they want from the query. There are a number of possibilities.

- Find objects similar to that specified by the query selection expression. This is a basic content-based retrieval (*CBR*) query.
- Find links defined from objects similar to this object. In this case the query selection expression will be matched against the selection expressions that are source anchors of links. This is a traditional multimedia content-based navigation (*CBN*) query.
- Find out which concepts, if any, are associated with this object. In this case the query selection expression will be matched against selection expressions which are lead-ins to concepts. Individual MMTs may be switched on and off, as the user might be interested only in the subject covered by a particular MMT or MMTs.

Once it is known which of these the user wants, it is simple to determine which selection expressions in the Selection Expression Layer should be matched against.

3.3.3 MMT Control Parameters

Even if the user does not want a concept from the MMT returned from the query, the MMT can be used to assist content-based navigation and retrieval. If the query selection expression is determined to be associated with a particular concept, a query can be expanded by searching with other representations of that concept from the MMT. In this way for example a link authored from an image showing one view of an object can be followed when the query matches with another view of the same object, or any other lead-in (of any media type) to the concept. If the query is a very general one, representations of related concepts in the MMT can also be used to expand the search. Use of the MMT to expand queries may be explicit or implicit.

The user can alter all of these parameters, though a sensible set of defaults will allow the user to perform queries quickly without the need for explicit specification.

The user may also want to control agent parameters, however these are outside the scope of this documentation.

3.4 The Matching Process

In order to evaluate a query and generate results MAVIS 2 must use content based matching methods to compare the query selection expression with selection expressions in the system. Each selection expression may refer to several selections and these may be qualified, indicating that only certain properties of the selection are important. Comparing two selection expressions can be achieved by comparing sets of selections on the basis of common properties.

To compare two selections using a property, a measurement of the property is required. There may be several ways to measure a property and each measurement from a particular algorithm is called a signature. Signatures will typically be calculated once and cached along with the selections from which they were derived.

Two signatures from the same algorithm can be compared to give a similarity measure, but these measures cannot be compared with those derived from another algorithm since the measurement scales are arbitrary and algorithm dependent.

One solution to this is to order results based on signature comparison and then merge lists of results (from several signatures and several properties) on the basis of ranking within the signature derived lists. The ranking within the lists is not dependent on the signature algorithms.

There are several approaches to merging these intermediate lists to produce a final list of results and priority can be given to matches based on quality or on the number of properties used.

3.5 Results

The results of a query will be presented to the user as a ranked list. An abbreviated representation of the matching selection expression will be shown together with the properties matched and an indication of the strength of the match. The user can specify exclusion thresholds to constrain the number of results shown, and can sort the list according to which properties are considered most important. To assist in the selection of thresholds, the distribution of match qualities of each property will be available.

From this list the user can view the destination of links (be they other selection expressions or concepts.) In addition the matched selection expressions can be used to modify the original query selection expression to express which of the selection expressions represent what the user is interested in and those which the user considers irrelevant. This is a form of relevance feedback, a technique firmly established as a useful tool in information retrieval [22, 45].

3.6 Prototype Implementation

A prototype user interface has been implemented using Java. A screenshot of this is shown in figure 3.2. A description of each of the windows follows.

3.6.1 MAVIS 2 Toolbar

The MAVIS 2 toolbar is where all MAVIS 2 operations start. The five icons on the left represent the core MAVIS 2 windows. The icon on the right is a *query shortcut.* Selections or selection expressions can be dragged onto this icon to perform a query quickly according to a set of user-defined defaults defined for that icon. The user can add any number of shortcuts.

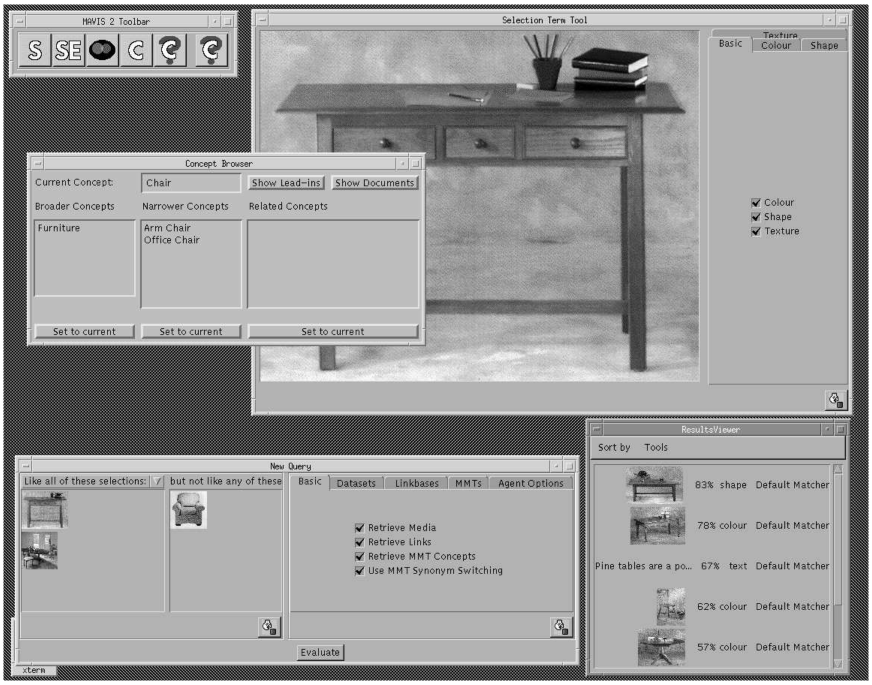

Figure 3.2: MAVIS 2 Prototype User Interface

These buttons start up one of the tools described below.

3.6.2 Selection Term Tool

When clicked on, the *S* button brings up a new *Selection Term Tool* window (as shown in figure 3.2) but with no media. Using this window, the user can describe an abstract selection, that is, a selection term (a term for a selection expression) with no actual media, whose properties can be defined by the user. In this way the user can specify an abstract property such as "red" or type in some text.

In addition this tool allows the user to load any piece of media as a selection term and select the relevant properties.

Terms from selection expressions can be edited by dragging them onto this button. This action opens a selection tool window with the dragged term shown and relevant properties shown on the right hand side of the window.

3.6.3 Selection Expression Tool

The button marked *SE* when clicked opens a new *Selection Expression Tool.* The left half of the "New Query" window shown in figure 3.2 consists of a Selection Expression

Tool. On the lefthand list of the tool can be dropped selection terms that are relevant to the query. Selection terms deemed irrelevant to the query can be dropped onto the righthand list.

3.6.4 Query Tool

The button marked with a small Venn diagram creates a new *Query Tool* window when clicked on. It is from this window that all queries are initiated (apart from those started from a query shortcut described above.) The selection expression tool on the left can be used to refine the query after it is first evaluated. The tabbed pane on the right is a *Scope Tool*, and allows the user to set the scope of the query, i.e. what type of results are required, how and which MMTs and/or datasets should be used, and what contribution agents should make.

Once the query has been evaluated the results are shown in a *Results Viewer*. This shows a ranked list of results which can be sorted according to the user's preference. Results can be clicked on to follow a link, show a link destination or source anchor, and to start browsing the MMT.

3.6.5 Concept Browser

The *C* button starts a new concept browser window. From this window MMTs can be browsed.

3.6.6 Query Shortcut Creation Button

The leftmost button marked with a question mark and a *C* allows the user to create a new query shortcut. It consists of a *Scope Tool* in which the user can set up options for a common type of query (for example, "which concepts in the MMT pertain to this object.") A new query shortcut button is added to the right of the toolbar.

Whenever a selection term or selection expression is dragged onto the shortcut the query is evaluated with the given options. A new Query Tool window is opened so that the user can refine the selection expression and other options if necessary.

4 Conclusions and Future Work

4.1 Summary of Completed Work

In summary, the following tasks have been accomplished:

- A detailed survey of relevant literature has been performed.
- A multimedia thesaurus structure has been designed, and possible uses of this structure identified.
- A number of prototypes of the concept browser component of the MMT has been implemented.
- MAVIS 2, an open hypermedia platform incorporating a multimedia thesaurus feature has been designed in collaboration with other members of the MAVIS 2 development team.
- A prototype user interface of MAVIS has been designed, identifying user-friendly ways in which the MMT (and other navigational techniques) can be used.

4.2 Evaluation of The Multimedia Thesaurus

The Multimedia Thesaurus offers a rich array of possibilities for aiding the user in locating their information need. The following possibilites have been identified.

1. Using only lists of representations (synonyms)
   (a) Expanding a query with additional representations
   (b) Switching the source anchors of links to a preferred form
2. Using a connected MMT
   (a) Expanding queries with representations from semantically related concepts

(b) Indexing media objects

(c) MMT navigation

(d) Narrowing the scope of a query to a portion of the MMT

3. Further techniques

(a) Using a richer set of semantic relationships

(b) Weighting relationships

(c) Constructing a search space

Testing the techniques 1a to 2d involves implementing the features, and designing an experiment to assess users' performance with respect to existing methods. This experiment is likely to take the form of providing the user with an interface and example dataset, and a set of tasks for them to perform (such as "find four Ming dynasty vases".)

4.3 Future Work

Multimedia information retrieval and navigation is a high-paced researched area. Hence new literature introducing new techniques and refining existing ones are constantly emerging. During the rest of this work it will be necessary to keep up with these changes. Thus the literature review is an ongoing work, constantly updated to reflect the current state of research in the field.

Before the multimedia thesaurus evaluation can proceed, the MAVIS 2 system must be implemented. This implementation will be a collaborative effort. During that time, detailed experiments will be designed to evaluate the MMT with as much emphasis put on quantative study as possible.

A time plan for the remainder of this work follows.

- Basic MAVIS 2 features implemented — end August '98
  - Inter-process messaging
  - Allowing content-based retrieval and navigation
  - Allowing viewing of at least text and images, and selections to be made
  - Logging facility

  - MAVIS design/implementation document

- Multimedia Thesaurus features implemented — mid September '98

  - MMT process design
  - Specification of interface with core of MAVIS 2
  - Concept browser/selector
  - MMT preferences tool
  - MMT logging facility
  - Add MMT documentation to MAVIS design/implementation document

- Multimedia Thesaurus testing — end September '98

  - Example dataset
  - Beta testing

- Multimedia Thesaurus evaluation — end January '98

  - Design and enter evaluation dataset
  - Design tutorial for evaluators
  - Design tasks for evaluators (e.g. "Find four Ming vases")
  - Design criteria for evaluating MMT
  - Perform tests with control group using classic CBR/N
  - Perform tests with group using MMT almost excusively
  - Perform tests with group with freedom to use MMT and CBR/N as they see fit
  - Collate results

- Results Interpretation — end February '98

  - Visualise results and evaluate how the MMT has been effective
  - Write up results

- Writing up — end August '98

- Complete MAVIS 2 implementation document
- Extend literature review
- Decide on thesis strucutre
- Write up MMT design
- Write up results and conclusions
- Write up possible future research

A Literature Review

Firstly, the precursor to the **MAVIS 2** system, **MAVIS 1**, is discussed. A summary of the features of traditional and electronic thesauri follows, and this section concludes with a survey of current and previous research of a similar vein.

A.1 The MAVIS 2 Proposal

The classic legacy hypermedia system has long been known to be deficient in many areas [25]. The **Microcosm** [19, 25] open hypermedia system addresses some of these deficiencies by allowing links to be authored on the content of a textual phrase as opposed to just a specific occurence of a phrase. This mechanism is known as the generic link.

The first **MAVIS** (Multimedia Architecture for Video, Image and Sound) was designed around the Microcosm architecture, and was an attempt to extend this functionality to allow links to be authored on the content of non-textual media[36]. However the approach taken by the first MAVIS project (described in detail in section A.2) does have some drawbacks:

1. If a generic link is authored on a particular view of an object, it may not be reached from a significantly different view of the same object.

2. If two destinations are relevant to an object, a separate link must be authored for each type of view to each of the destinations [5], and the semantic equivalence of these links is not expressed.

The **MAVIS 2** proposal [26, 27] aims to overcome these problems by providing better navigational support. The proposal to the **EPSRC**[27] details the following four objectives:

1. Identify and evaluate methods for specifying and recognising objects in media

2. Design a multimedia thesaurus extension to the distributed open hypermedia architecture
3. Investigate how intelligent agents can enhance retrieval and navigation in this architecture
4. Evaluate these enhancements in terms of author and end user productivity.

To fulfil these objectives three extensions to the MAVIS architecture are proposed:

- The Multimedia Thesaurus (**MMT**)
- Intelligent agent support
- Model-based representations

Each of these extensions are described in the following sections.

A.1.1 The Multimedia Thesaurus

One way of solving problem 1 listed above is to provide a digital thesaurus facility.

The key features of the MMT are described in the MAVIS 2 proposal[27]:

1. When a link is authored from a representation in the MMT, it will be possible to follow this link from any other representation of the same object in the MMT.
2. It will also be possible to follow links for semantically related objects.
3. A content-based retrieval operation can retrieve not only directly matching objects but different views of those objects too, and these different views may be used to widen the scope of the retrieval.
4. The information held in the MMT could be used to assist the user when initially entering media, for example by indentifying objects in the media already held in the MMT.

The description of feature 1 makes an implicit assumuption about the architecture of the MMT — it assumes that representations of objects are held within the MMT, or are at least considered part of the MMT.

Paul Lewis describes a possible MMT in further detail [36]. Firstly a hypothetical textual thesaurus for enhancing textual generic links is described.

The thesaurus could take the form of a network of terms related by the relations *equivalent term*, *preferred term*, *related term*, *broader term* and *narrower term*. Whenever a link is authored from a term, the source anchor is switched to the preferred form of that term. When links are requested from a term, the linkbase is searched for the preferred form of that term. Thus the problem of different terms representing the same concept is eliminated. These operations could be performed invisibly, and could be implemented as a filter in the the filter chain of the current Microcosm architecture.

In addition, during link following the user could widen the scope of the search by "fanning-out" the search to include related or broader/narrower terms.

The same paper also discusses possible multimedia extensions to this somewhat conventional architecture. Instead of consisting of terms the thesaurus consists of *media-based representations* (in this case "representations" refers to relevant signatures derived from the media) of any media. Links authored from one media may be followed from representations in other media. The preferred representation should be textual however for ease of searching the linkbase.

Another "natural" extension of this architecture is also described. A thesaurus stripped of equivalent representations can be considered a conceptual network, consisting of just preferred textual representations and the relationships broader representation, narrower representation and related representation.

Other types of semantic relationships can be added providing a richer way of broadening, narrowing or even shifting the scope of the search for links. Specifying this could be achieved by explicit user navigation of the preferred representations.

Lewis raises some other issues:

- Will the representations (signatures) be held within the MMT, and will the selections from which they were derived be held within the MMT?

- During MMT browsing, how will the system indicate the existence of non-text representations and their relationships?

- Should MAVIS 2 be built around the MAVIS/Microcosm architecture?

These are all important issues to be considered during the design of the MMT architecture.

A.1.2 Intelligent Agent Support

Another aspect of the MAVIS 2 project is the integration of intelligent agents to aid navigation and retrieval, and to seek new semantic information using the relationships and representations held in/associated with the MMT[26, 27].

Two types of agent are identified in the proposal [27]:

Descriminator Uses neural net and other techniques together with knowledge from the MMT to find new ways of discriminating between objects of different classes.

Classifier Seeks clusterings in the "accumulated knowledge base" (the MMT and associated representations). Should provide richer semantic associations for link following.

The proposed use of agents in MAVIS 2 has been somewhat revised since then. In an early internal paper[34], Dan Joyce describes a multi-order agent architecture. Simple first order agents with simple goals traverse media and the MMT until their goal is achieved. Second order agents monitor and interpret the dynamic and static state of these agents, and a possible third order agent collaborates these results and presents them to the user.

These agents could work in a distributed environment using some mobile agent architecture such as that advocated by Jonathan Dale[15].

While this area of research is not my key concern, it is part of the MAVIS 2 project and the Multimedia Thesaurus must accomodate and facilitate the use of intelligent agents.

A.1.3 Model Based Representations

The proposal to EPSRC[27] mentions the investigation of model-based representation techniques. The MMT architecture must be able to accomodate arbitrary representations so the investigation of this area is being postponed until the MMT architecture is decided.

A.2 The MAVIS 1 Project

MAVIS 1 is an attempt to provide *multimedia* generic link capability to the Microcosm open hypermedia system [36]. In the original Microcosm system [25], generic links are created by selecting a piece of text and a destination document. This link can be followed by selecting any occurrence of that piece of text and invoking a follow link action. MAVIS 1 attempts to allow this functionality for other types of media, such as images, video and sound.

The core architecture of MAVIS 1 (shown in figure A.1) is modular, the actual matching methods used are implemented as "signature modules." These are media-specific matching modules that each use a different method to determine the similarity between two pieced of media. This approach is necessary because of the *multimedia matching problem* described in detail in section A.4.1. Briefly, the two main advantages are:

- It allows future amendment and addition of media matching methods.
- The user can select relevant matching methods for task. For example the user may wish to find a link on an image using its colour.

In MAVIS 1 the user can select a piece of media and have it offer a number of links for following (content-based *navigation*), and in addition can elect to find all pieces of media in the Microcosm document management system [19] matching the selected piece of media (content-based *retrieval*). Note that the two processes are very similar, the difference lies only in the scope, i.e. which pieces of media are candidates, source anchors of links or everything in the system.

A.3 Thesaurus Systems

Broadly speaking, a thesaurus is a set of terms connected by a set of relations[33]. Broadly speaking there are two types of thesaurus — those constructed automatically, and those constructed manually [31]. Manual thesauri are detailed first.

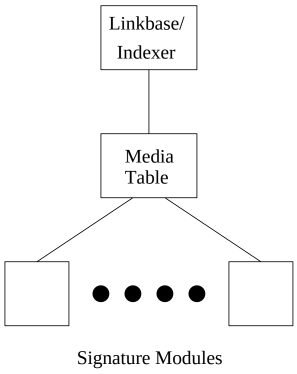

Figure A.1: MAVIS 1 Architecture

A.3.1 Manual Thesauri

Manual thesauri are those constructed by authors as opposed to any automated electronic process. The category can be broken down further into two sub-categories. The first is the word-based, general purpose thesaurus, such as Roget's thesaurus. These are largely used to find alternative or preferred forms of a single word or term. The second is the phrase-based, retrieval-oriented thesaurus such as the INSPEC classification thesaurus [39], or the Arts and Architecture thesaurus [3].

The former type of thesaurus is largely in hand-held book format or an accessory to a word processing application. It's only use in an information retrieval environment is query expansion (supplementing a query with synonyms of the original query term.)

The latter is commonly used in IR environments to index large collections of documents. Since this is essentially the task of the MMT[26, 27, 37] documentation and research in this area is likely to be very relevant to this work.

Structure/Relations The majority of thesauri use five main relations (or close equivalents):

- **Broader term** The *broader term* of term A is a term encompassing the scope of term A.
- **Narrower term** The complementary relationship of *broader term* - the scope of a *narrower term* B of term A is encompassed by the scope of the term A. If this relationship exists, term A is always the broader term of term B[1].
- **Equivalent term** Term A is *equivalent* to term B if term A is synonymous with term B. In a thesaurus with a relatively low level of detail (specificity),

term A may in fact be broader or narrower than term B, if the difference in specificity is small enough.

Preferred term This relation is very much like *equivalent term*, however this relation is used to indicate the preferred (often the most common) form of the term.

Related term Used when term A and term B have some semantic association, but their relationship cannot be adequately expressed in terms of one of the above relations.

If the thesaurus as a whole is considered as a concept map, the *related term* relationship can be considered as a mapping for all non broader/narrower relationships[31].

The above relations may be given weightings, depending on the strength of the connection. This is particularly useful for *related term*, where information about the connection is very limited.

A thesaurus, then, can be considered a map of some subject domain [33]. There are a number of different ways of visualising this map:

Arrowgraph A graphical representation, with physical distance between terms on the diagram related to "semantic distance" (how closely the terms are related.) These diagrams are difficult and time-consuming to produce and maintain, although methods are emerging for automatic generation of these or similar diagrams [21].

Flat Map Essentially the thesaurus is a collection of nodes (terms) and links (relations) with no structural restrictions. Thus it can be considered simply a flat concept map. This approach offers no advantage and does not aid navigation of the map in any way [1].

Hierarchy By far the most common visualisation, this approach assumes that the most useful relationships in the thesaurus are of the broader/ narrower type. The terms are arranged as a hierarchy, moving up the hierarchy involving traversing broader term relationships, moving down the hierarchy traversing narrower term relationships. Typically the thesaurus has a number of broad subject headings (known as *facets*), below each of these is a hierarchy of terms encompassed by this heading. A set of equivalent/ preferred terms are considered as single nodes in the hierarchy, or commonly only the preferred form of a term appears. Although because of related term connections the hierarchy is not strict, the hierarchical core of the thesaurus is used as the basis for visualisation.

Specificity Refers to how detailed the terms in the thesaurus are[1]. A term or part of a thesaurus with high specificity covers a relatively narrow subject field. A term or part of a thesaurus with low specificity covers a relatively broad subject field.

Recall A thesaurus which, when given a search term, will offer a (relatively) large number of documents is known as one with good *recall*[33]. This is only a measure of the how many documents are offered, and not of the relevance of these documents. This terminology is also applied to content-based retrieval systems.

Precision The complement of *recall.* The precision of a thesaurus is the proportion of retrieved documents *relevant to the user's query.*.

Controlled/Free Language In a controlled language thesaurus, only a controlled subset of all terms in the thesaurus are actually used to index documents or represent the relationships between concepts[1]. Other terms in the thesaurus "lead-in" to the preferred term for a particular concept.

A controlled language thesaurus aids retrieval at the (possible) cost of specificity. This means that given a search term, a controlled language thesaurus is likely to retrieve a number of documents, though the chances of each of these documents being relevant to the searcher is reduced.

Faceted Classification Each term in a thesaurus may also appear (usually hierarchically) beneath one of a group of *facets* [52]. Each facet reflects one aspect of the overall subject being covered. A facet may simply describe the fundamental type of the term, for example "Process" or "Material," or may describe the term in more detail. For example in a thesaurus describing chemical manufacture, there may be facets "Raw Materials," "Catalysts," "End Products," "Apparatus," and

"Processes." For a more general thesaurus, a more generic classification schema may be adopted such as Ranganathan's *PMEST* formula [41] ("Personality," "Matter," "Energy," "Space" and "Time.") However, in all cases this classification schema is predetermined and considered too time-consuming to alter once the initial thesaurus has been constructed [52].

By far the most common form of thesaurus is a controlled language, faceted hierarchical thesaurus. Other structures exist such as multitrees[20] but these tend to be suited to particular applications.

A.3.2 Automatic Thesauri

Automatic generation of thesauri is not a new field, research in the automatic classification of terms goes back as far as the 1960s and earlier [44, 32]. However with the advent of widely used electronic Information Retrieval systems and high-power computing, far more complex methods can be used prompting more recent research in the area.

A number of methods exist to automatically create thesauri (or similar semantic structures.) All investigated are text-based, using text analysis techniques to identify candidate terms [14, 31]. An iterative process known as relevance feedback, involving analysing the context of a term in many different documents, can be used to refine this process [31, 22].

These methods may prove useful as a method of constructing an initial thesaurus for an application, but are outside the scope of this research.

A.3.3 The Okapi Projects

The **Okapi** experimental system is a text retrieval system which is used as a basis for evaluating the effectiveness of retrieval techniques [42]. The overall structure of the system is shown in figure A.2. Although the architecture is oriented to retrieval as opposed to a hypermedia style of navigation, it can be seen that the thesaurus is not used as an indexing tool, but merely as a resource for the query model.

Explicit and implicit use of the thesaurus has been tested[4, 43], and in some of the experimental user interfaces thesaurus navigation has been available.

Most interfaces to Okapi are designed for non-expert searchers. The question as to whether MAVIS 2 should be designed with non-experts in mind has not been explicitly

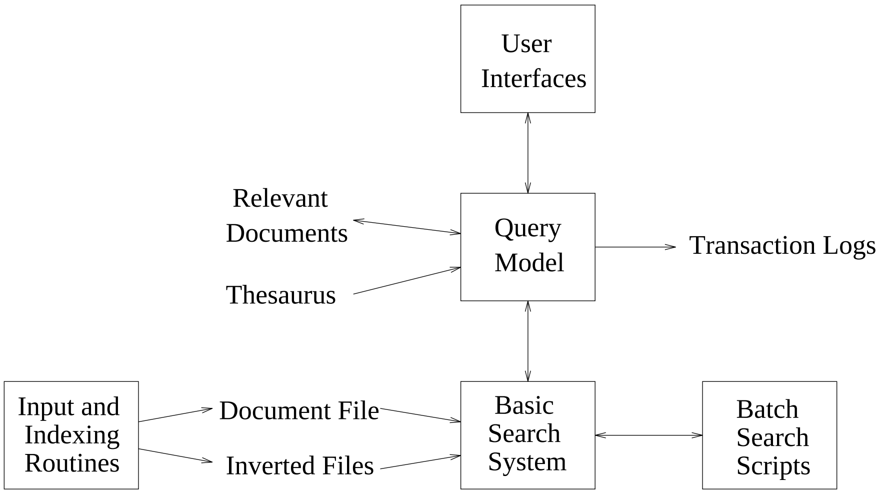

Figure A.2: Structure of the Okapi System

explored, however to exclude non-expert searchers as potentional users would seem unwise. The basic process of an Okapi query is as follows:

1. The user enters a free-text query.

2. The word stems are extracted and given a weights based on their frequency in the document collection.

3. The system produces a ranked list of documents, showing the user the top scoring titles.

4. The user can view any of these documents, and upon doing so is asked "is this relevant?" to which the users answers "yes" or "no."

5. The system extracts terms from the documents marked as relevant and constructs a new query. This is referred to as *query expansion*, although some terms may be dropped from the previous query. This is a form of *relevance feedback* (see section A.4.3).

6. The process can then iterate from step 2 above. The user may or may not be given the option of manually including or excluding terms from the query.

The thesaurus may be used in different ways. In a paper detailing experiments with user interfaces [4], two different methods of thesaurus use are mentioned.

The first involves explicit use of the thesaurus. Given the terms of a query, synonyms and closely related terms from the thesaurus are presented to the user, which the user may add to the query. These terms may also be used to enter and navigate the thesaurus, from which additional terms may be added to the query. No details concerning the thesaurus navigation process are given in the paper.

The MMT in MAVIS 2 could be used in this manner to expand a query for content-based retrieval. Such a process could be used to facilitate a search of some information not already associated with concepts in the thesaurus, such as World Wide Web documents.

Implicit use of the thesaurus takes the form of matching terms in the thesaurus and using synonyms (and maybe connected terms) to expand the query, *without* interaction with the user. However the paper does not detail the relative benefits or shortcomings of this approach as opposed to the method described above.

The paper describes both methods as being useful, though the user's selection of relevant terms from the thesaurus was noted as being especially valuable. This could have implications for the uses of the thesaurus described in the proposal[27] whereby the thesaurus is used to "flesh-out" a query for content-based retrieval. In such a process an interactive method of picking which media objects to use for this retrieval may be useful.

A.4 Content-based Retrieval Systems

A.4.1 The Matching Problem

One of the principle problems encountered in multimedia information retrieval and navigation systems is the *multimedia matching problem.* "Matching" refers to the process of determining how similar two pieces of media are, with a view to determining whether or not the two media objects represent the same real world object[24].

Text is a relatively straightforward medium to match. It is concise, and can be discretely represented in a computer with no loss of information. Given two pieces of text, it is trivial for a computer to work out whether or not they are the same; it need only compare the ASCII codes.

Such a straightforward technique can be supplemented in many ways to recognise different forms of words. There are many algorithms for word "stemming" — comparing the stems of words allows the recognising of different tenses etc. of the same word[46]. Some systems use simple thesauri (merely lists of synonyms) to identify different forms of words.

Testing for similarity in other media is not so trivial. In the case of video, image and sound, two media objects representing the same real world object will not be exactly the same, even in the difference is unobvious to the human eye or ear. There are a number of problems associated with matching non-textual media.

A.4.1.1 Similarity Measures

There are many ways in which two pieces of media can be considered similar. In the case of images, they may be similar in colour or shape. In the case of video, they may depict objects moving in similar manners. Two sounds may be of a similar pitch. Each such characteristic of a media type is known as a *property* or *feature* of that media type.

Not only can each media type have a number of features, but different methods exist for measuring and comparing these features. For example, there are many ways of comparing the colours of two images, such as RGB histograms or spatial colour distribution.

Content-based retrieval systems attempt to address this problem by using algorithms to estimate the *similarity* of two pieces of media [30, 48, 2]. Typically, given two media objects (of appropriate type), an algorithm can give some mathematical indication of how similar the two objects are in terms of a property.

Retrieval systems usually require the user to provide an example media object and specify a property or properties of the example. The system then uses the results of applying the algorithm to this example to find similar objects from an index. This produces a ranked (according to the results of the algorithms, and any weightings provided by the user) list of objects in the system. However since the result is fuzzy and not boolean, it is likely that *every* object of the same media type as the query in the system matches the query to some extent. Thus the list is usually thresholded, either by only considering the top n matches, or selecting a cut-off value for the match strength.

There is a wealth of research in the field of media matching technology [9, 7, 2, 30, 48]. Most is application specific, and some is geared towards finding quicker ways of finding a similar object from a large database of candidates. Some or all may be applicable in some MMT application. The MMT should be designed with a view to allowing potentially any such matching algorithm to be applied.

A.4.1.2 Different Views of an Object

Even though two pieces of media may represent the same object, the media may be significantly different. For example, two images might consist of one object viewed from different camera angles, or two sounds might be of different voices saying the same thing.

A handful of algorithms attempt to address this problem. A three dimensional model of an object can be used to identify an object at any orientation. This method is however a fledgling one, only works for relatively simple objects, and is computationally expensive[50].

There is also the question as to whether it is *appropriate* to offer a media object depicting a different view. The user may want to find similar *media objects* to the query, or to find media objects *depicting the same real world object(s)* as the query. In the case of the former, the problem of different views is not applicable. In the case of the latter, some other technique must be used.

Existing techniques rely on producing some representation of the real world object without media examples, such as the three dimensional model method. An alternative approach is to store some metadata indicating which media objects represent the same real world object. A good match with a media object then indicates that any of the associated media objects may also be relevant to the query [49].

A.4.1.3 Recognising Objects in a Larger Context

Most content-based retrieval systems will only retrieve a media object that matches well (using whatever similarity measures are appropriate) in its entirety. Consider a query consisting of a small media object. This query object may appear in objects in the system, though that object as a whole might not match well with the query object. For example consider a query consisting of an image of a car; the user may wish to retrieve scenes or videos *containing* cars, though the scene or video as a whole might not produce a strong match with the query object. The scene will not be retrieved unless the car is entered into the system as a media object in its own right.

There are methods for locating objects in scenes such as *template matching* [50]. These methods are usually computationally expensive. Currently, some (possibly minimal) form of user interaction is usually required when selecting objects in a scene.

A.4.2 Retrieval Models

Retrieval systems come in a variety of flavours [46]:

Boolean In such a system documents that contain all terms in a query are considered relevant, and all others irrelevant. This is the most common type, especially in commercial text retrieval systems[40]. Due to the necessity of being able to determine for certain whether or not a document contains a term, this technique is unapplicable to the multimedia domain without some modification.

Vector-based Documents and queries are represented as vectors of dimension n where n is the number terms in the document space. Each element corresponds to a weight. Some measure is used to estimate the similarity of documents and queries, usually the cosine of the angle between the vectors.

Probabilistic The probability of a document's relevance to a query $P(R|d_i q)$ is estimated using Bayes theorem. This produces a ranked list. The *INQUERY* retrieval system uses a form of this technique[10].

Feature Space To address the matching problem described in section A.4.1, a derivative of these models, the *feature space* has been developed[50]. This technique is still only capable of retrieving media of the same type as the query. Each measurable feature of a query or document is a dimension in the feature space, thus each object has a coordinate. The euclidean distance between the query and document is the measure of the similarity of the query to the object.

The feature space can be used to *classify* objects as well. If we make the assumption that objects of a certain class (e.g. landscape scenes) will have similar features, the objects of that type will be close together in the feature space. Over a large collection of objects clusters of similar objects will form. Classification of query objects can be attempted according to which cluster it is closest.

A.4.3 Relevance Feedback

Relevance feedback is a long-established technique in the text information retrieval world [45, 55]. It is the process wherby once the initial results of a query are obtained, the user indicates to the system those retrieved objects that interest them, and those that do not. These are used to *expand* or *refine* the query and produce a new set of results that should better meet the user's infomation need.

A.4.4 Existing Systems

Several database systems already exist allowing the retrieval of non-text media. Most allow the assigning of keywords to images too, however this often requires excessive authoring effort in large data sets[40].

QBIC The Query by Image Content (*QBIC*) system developed by IBM[18]

Virage Inc.'s Visual Information Retrieval Tool is an extension (called a "datablade") for Informix Software Inc.'s DBMS[53].

VisualSEEk and Webseek A region-based CBR system for images [47, 54].

Excalibur Technology's Visual RetrievalWare [17]

MARS The *Multimedia Analysis and Retrieval System* researched by Ortega et al. [40] uses a probabilistic boolean retrieval model.

A.5 Other Approaches

This section describes previous and ongoing research of a similar vein to the MMT.

A.5.1 The Himotoki System/COIR

A hypermedia paradigm with possibly the closest goals to those of the MMT and MAVIS 2 is currently being researched by Hirata et al. [28, 29] at the C & C Research Laboratories, NEC USA Inc.

Their Hypertext 96 paper [28] describes the concept of "content-oriented integration." However the English in the paper is of poor quality as is the structure. Many assumptions must be made when interpreting each paragraph.

For example, “each media representation [held in the system] has the part which can be translated into the conceptual representation.” This may either mean the representation is associated with a corresponding conceptual representation, or that e.g. a keyword is held with each media representation identifying the conceptual representation to which it is associated. The term “representation” could mean an actual piece of text, or image etc., or a derived feature description (like a signature.) In addition it is not made clear (anywhere in the paper) whether each conceptual representation has a single media representation or can have many media representations. These points are not clarified elsewhere in the text.

From careful reading of the paper, the basic concept of content-oriented integration can be described thus: Each object for which there exists a piece or pieces of media has a conceptual representation. This conceptual representation is bidirectionally linked to each piece of associated media. The user can navigate either using purely media, or may choose to navigate via the conceptual representations.

The approach has the following features:

- Using conceptual navigation, navigation between different media types is possible. (The MAVIS 2 MMT should provide this functionality.)

- Semantic relationships between objects are stored between conceptual representations. This allows the relationships to be used for navigation independent of media type. (This is also true of the MAVIS 2 MMT.)

- Classic, media-type-dependant, content-based retrieval can take place independent of conceptual representations. (The MAVIS 2 architecture should allow for this — if the existing MAVIS/Microcosm architecture is used this functionality will already exist.)

- Media-based clues and conceptual-based clues can be combined. For example, a video in which a butterfly moves from the left to the right of the screen could be retrieved. (Such functionality is an interesting research area. How this is achieved is not well described in the paper, however it may be possible to provide this feature with the MMT.)

- A modular design allows media-specific processing methods to be changed, inde-

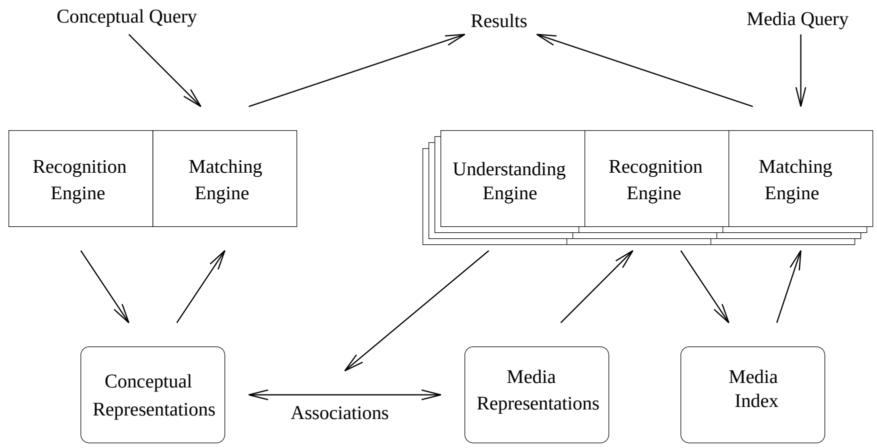

Figure A.3: Hirata's Content-Oriented Integration Architecture

pendently of other media types and the conceptual representations. (The media-table style architecture of MAVIS 1 allows this, as should MAVIS 2.)

The architecture proposed by Hirata's paper has many of the same goals as MAVIS 2. It is quite a generic architecture, with modular parts enabling for example new media processing and understanding technology to be supported. An overview of the architecture is shown in figure A.3.

Media and conceptual representations are dealt with by the media and conceptual *augmenters* respectively. Each augmenter has a *recognition engine* and a *matching engine*. In addition media augmenters have an *understanding engine* which translates a media representation into a conceptual one. Since conceptual representations are already "understood" and associated with media representation(s) the conceptual augmenter does not require this engine.

The features of each of these engines are as follows:

Recognition engine The media recognition engine extracts media attributes and stores them in a media index. This may consist of many different modules for different recognition algorithms. The conceptual recognition engine indexes the conceptual representations.

Understanding engine This engine uses the media attributes extracted by the recognition engine to attempt to find a corresponding conceptual representation for the media representation.

Matching engine This object, when given a query, finds the matching media or conceptual representation. The media matching engine also employs **media-based clustering** techniques. If the query is far from a cluster then that cluster is elminiated from the search.

This architecture allows for different types of navigation:

Media-based navigation This encompasses content-based navigation and retrieval using the media augmenter. This type of navigation is available in MAVIS 1. Additionally, the user may provide media clues as a query, for example by sketching an outline. A simple extension to the MAVIS 1 architecture could allow that style of navigation.

Conceptual-based navigation As well as allowing navigation around conceptual representations (the representation itself is textual in form and thus can be displayed,) media to media navigation via conceptual representations can be achieved by using the understanding engine of the media augmenter to identify the relevant conceptual representation, and another media representation reached from that or another associated conceptual representation. This seems analogous with navigating with the MMT.

Schema Navigation The description of this style in the paper is very brief and sketchy. Both conceptual-based and media-based schema navigation are possible. Conceptual-based schema navigation is apparently where the system designer defines the conceptual representations and relationships around which the media representations are linked. The only difference between this style of navigation and conceptual-based navigation above would appear to be that the conceptual representations are author-controlled and not created by some automatic algorithmic approach. (How the MMT is constructed is also an area of interest — an architecture could allow either method.)

The paper also mentions media-based schema navigation. This is poorly explained. It could include guided tour styles of navigation and the traditional hypermedia explicit point-to-point links.

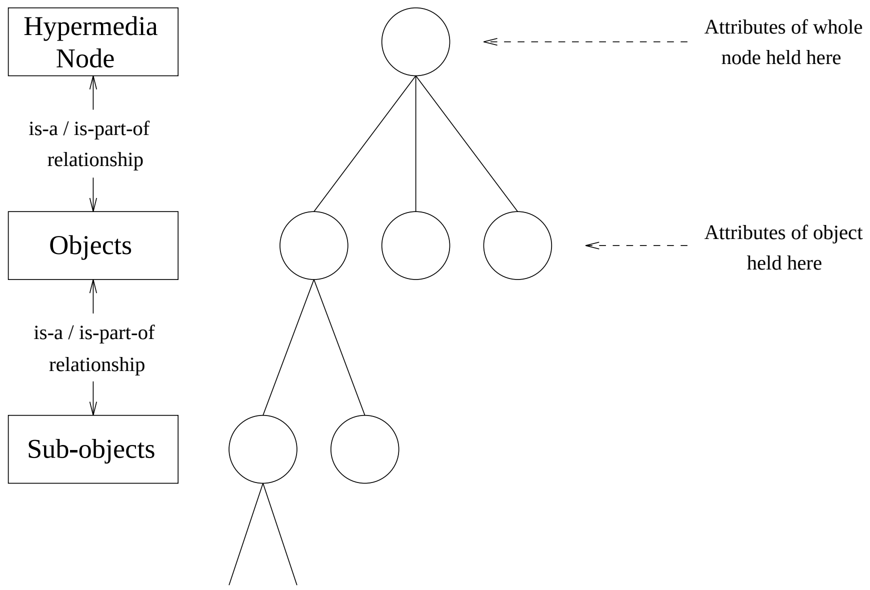

Figure A.4: Hierarchical Attribute Structure

The **Himotoki** implementation of this architecture is described in the paper. Himotoki is a client-server style architecture. The client merely handles the GUI aspects, all the query processing and navigation operations are handled by a server.

The second of Hirata's Hypertext papers[29] is written in slightly better English. It describes an extension to the content-oriented integration architecture described in the previous paper called *object-based navigation*, however it is re-presented as a generic set of hypermedia tools called the *Content-Oriented Information Retrieval* tools (**COIR**).

Multimedia data in the system is broken down into *objects*. The user can form a query specifying multiple objects and the relationship between them. To allow this feature the attribute data of each piece of media is hierarchically structed. This structure is shown in figure A.4.

Each hypermedia node (minimum unit of navigation) is split up into objects. Each of these objects can be further sub-divided. This decomposition is stored as a tree in which each node corresponds to one object or sub-object. The root of the tree represents the whole hypermedia node.

At each node of the tree (including the leaves) are held the attributes of that object or sub-object. These include media-based attributes, and a possible association with a conceptual representation. In addition, the physical relationships between the direct children (sub-objects) are held, such as positional information from an image. The logical relationships between each object/sub-object are held with the linking branches. These are bidirectional is-a/is-part-of relationships. (While the MMT is specified as having only thesaurus-type relationships, it may be useful to expand these to include a richer set of semantic relationships including is-a/is-part-of type relationships.)

Structural constraint (specifying relevant nodes in the tree) and attribute constraint (specifying relevant attributes at each node) can be applied when matching to ensure the relevant qualities of the media/object are matched against. This is equivalent to the selection of signature modules to use in MAVIS 1[36].

During a query, users can specify a number of objects and relationships between them. Since this can be a complex operation involving the choosing of relevant media-based attributes, default settings can be used allowing navigation to be achieved using just mouse clicks. The query handler extracts the relevant attributes from the objects. How this is achieved is not explained in any detail.

The matching phase then invoked is divided into four steps. This process is very poorly explained and is full of ambiguities and gaps.

1. The initial phase described in the paper as “Correspond the objects with each other.” It shall be assumed that this step involves finding objects from the media index that match those of the query.

2. “Evaluates the similarity between each of the correlated objects.” This is taken to mean the strength of each match is calculated using media and conceptual attributes. These matches may be weighted by the user or application.

3. “Evaluates the similarity of the relationships.” Where strong matches are found, evaluate the similarity of the relationships between them (both physical and logical.)

4. “Evaluates the similarity between the query and the media index.” This can only be assumed to collate the results of steps 1–3 above and return the results.

While the explanation of the object-based query-match mechanism is very weak, it is a potentially very powerful mechanism. No similar mechanism is mentioned in any of the MAVIS/MAVIS 2 literature, however allowances for a similar mechanism in MAVIS 2 could greatly improve its flexibility and should be strongly considered.

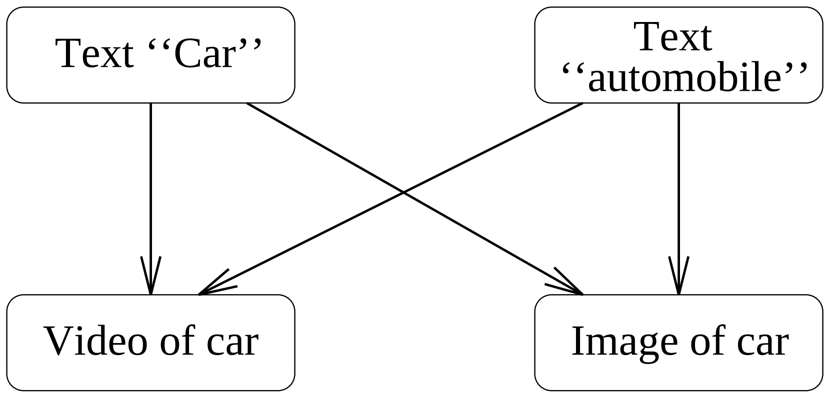

Figure A.5: Separate Links Expressing Same Relationship

A.5.2 Link Clustering

While some systems including MAVIS 1 allow for content-based navigation, they still have two main disadvantages that MAVIS 2 should address:

- The underlying semantics are not represented and thus cannot be used as a means of navigation.
- Two links which effectively refer to the same relationship are held separately, and their similarity not expressed to the user. For example in figure A.5, four separate links are expressed, all representing the same relationship. As the number of relationships and representations increase, the number of links required increases exponentially [5].

These problem can be alleviated by the use of link clustering [5, 13]. In figure A.6 the same information is stored. The central abstract representation "car" can be used to simply express relationships with other abstract entities, reducing the need for additional media to media links.

The central node of these clusters correspond to the conceptual representations of Hirata's content-oriented integration [28, 29]. They can also be considered as corresponding to the preferred representations in the MMT, where each of the source anchors is an equivalent representation, and the documents linked from the cluster are destinations of links authored on the representations. Associating all links from the central node of the cluster does however seem a more elegant approach as the searcher (be it the user or some automatic algorithm) need only navigate from the source anchor

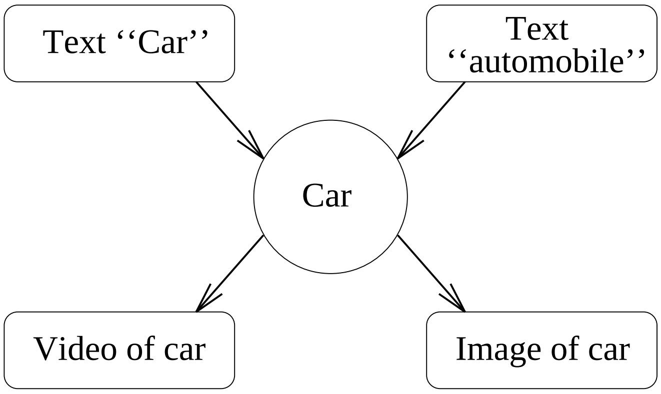

Figure A.6: Clustered Links

to the cluster and then choose a destination or destinations, as opposed to searching the linkbase for any links authored on each of the representations.

A.5.3 Semantic Indexing

Semantic indexing is similar to Hirata's object-oriented system described above, in that documents in the system are indexed conceptually and connected via semantic relationships, as opposed to being indexed by terms alone.

In an architecture proposed by Cunliffe et al. information items in the media base are indexed as queriable relationships [14]. These are expressed as binary relationships in the form:

< *OBJECT RELATIONSHIP SUBJECT* >

RELATIONSHIP is one of a finite set of relationships allowed by the system such as *A-Kind-Of*. These include hiearchical specialisation / generalisation relationships, so these relationships can be used to form a graph-style network of the knowledge base with a hierarchical core.

The database of relationships can be queried based on the *Standard Associative Form* which allows queries such as:

< *? A-Kind-Of Community-Life* >

These relationships are also weighted according to the strength of the relationship, and the specialisation of the subject and object. The reason for the latter is that two very specialised concepts close together in the network are likely to be more similar than two more general concepts.

Each of the concepts is associated with retrievable media items. The relevance of each media item can be assessed by the weightings attached to relationships, and the number of relationships traversed between concepts.

This architecture is similar in many respects to the MMT, with a richer set of relationships. However the MMT could be extended to allow richer semantics, and the idea of weighting relationships could be usefully applied, especially the idea of weighting affected by the relative specialisation of concepts.

In Cunliffe's example application, the relationships held are predetermined. However methods exist for automated semantic indexing, and these techniques may be applicable for the construction of an MMT.

Dumais et al. [16] describes a *Latent Semantic Indexing* approach. As opposed to indexing documents by individual words or terms, statistical techniques are used to estimate the amount of "noise" in this term to document relationship caused by varying word use. This involves a mathematical approach producing a term-document matrix. Each document title and term has a coordinate in this space, and the distance between titles and terms can be used to estimate the relevance of a term to a title, or of two titles.

These approaches are still text-based; applications of these for generic media would be difficult considering the impossibility of exact matching, and the inability to extract discrete objects from some types of media. For example images cannot be exactly matched and extracting the pure representation of exactly one concept is extremely difficult. However these techniques may prove useful and could be used during initial construction of the underlying concept network of the MMT. In addition, if text is available for all concepts held in the MMT such techniques could be usefully applied to improve retrieval performance.

A.5.4 The VISAR System

Most hypermedia and information retrieval systems, including those described so far, will retrieve media objects as opposed to conceptual ones. Hirata's object-based navigation does make use of conceptual representation in the querying process[29] but still returns media objects and not conceptual representations. The **VISAR** system presented by Clitherow et. al [11] attempts to improve retrieval by finding networks of concepts from a semantic net similar to that expressed in a query as opposed to media items. The user can then reach documents indexed by these concepts. Although only documents' textual titles are considered this represents quite a fundementally different querying philosophy — the whole querying process is concerned with the meaning of the query as opposed to the media itself.

The system has available a potentially huge collection of documents. A semantic network is formed by using natural language processing (the "Lucy" system) — basically it is used to deduce "who is doing what to whom." The resulting index is similar to the proposed MMT, with a richer set of possible relationships and mich finer granularity. It is implicit from the paper that the network is hierarchical in nature.

At any given time the user has a "focus of attention" — this is a small (and simplified) portion of the network that the user is interested in, with weightings assigned to each of the entities, which number no more than seven. (This is of course George Miller's "magic number"[38] — the capacity of the human short term memory.) This network is known as the user's default *perinrep* — *personal information representation*, the initial state of which is decided using a number of factors including the user's occupation and social background. No analogous features are described in the MAVIS 2 proposal — the user's "focus of attention" is considered to be the piece(s) of media or thesaurus concept representation(s) currently displayed on the screen. Neither does the proposed system treat individual users differently — no record of the user's interests is held.

The querying process starts with the entry of a textual query. Natural language processing is used to parse the query into conceptual units which must exist in the overall semantic network. Relationships between these units are inferred and a small network of concepts results. This is combined with the user's default perinrep to produce the request perinrep. A technique called *highest common concept* (*HCC*) is used to achieve this.

The HCC process is used to combine two sub-networks of a semantic network. This may have applications within MAVIS 2. This is especially true since the VISAR semantic network has a specialisation/generalisation relationship, which directly corresponds to the broader/narrower representation pair in a classic thesaurus.

Three reduction operators are used in the process:

Concept generalisation finds a common "ancestor" of two concepts — that is find a concept of which both operand concepts are a specialisation.

Concept factoring Simply joining two graphs with common concepts together.

Concept chaining is "the process of relating concepts through slots other than the generalisation/specialisation slot." This is most likely the process of merging two concepts between which a relevant relationship exists. Currently the slot is chosen from a list of "appropriate" slots gathered heuristically.

The reduction process consists of the following steps:

1. Construct the set L of all (unique) concepts from the sub-networks.

2. Sort these according to specificity — effectively semantic distance from the root of the overall network. The relative weightings of each element are taken into account.

3. Apply the HCC function on the last two elements of L replacing them with the result of this function. Repeat until the length of L is $<= 7$.

4. Define links between the n concepts in L based on their relationships (or paths of relationships) in the overall network.

The request perinrep can then become the user's default perinrep (focus of attention.) The user is free to expand/collapse concepts and navigate through the overall network, or ask for some relevant documents.

When documents are requested, the VISAR proceeds as follows: Using the concepts explicitly specified in the query first, the concepts in the request perinrep are used to perform keyword searches of the documents held in the system. Parts of the request perinrep connected to those concepts are then used. This process is repeated, getting further away from the originally specified concepts, until a terminating condition is reached:

- A predetermined number of citations have been found

- A predetermined amount of effort has been spent (e.g. CPU usage)

- The concepts found fall below a given threshold of relevance (semantic distance is greater than a threshold.)

The prototype VISAR system only holds titles but the whole document could be held and reached from these. The results yielded by this process have a high chance of being relevant to the user even if the wording is significantly different to that of the query. The aims of the system, then, are very similar to those of the MMT in MAVIS 2. The system only recognises text, and uses only the title of documents and not the content. In addition the querying process relies on the quality of the natural language processing technology used. However the returned documents are often relevant to the query, even though the wording of the title is significantly different to the extent that there are no terms synonymous with those in the query. The system also considers the relationship between the terms which the MMT as proposed does not. However the similarity between the way in which knowledge is held does allow for the possibility of application of some of the techniques in MAVIS 2.

A.5.5 Document Clustering

Thesauri (including the proposed MMT) and other knowledge storing networks are usually composed of atomic terms or concepts, and the knowledge is represented with the relationships between these. An alternative structure is the Vector Space Model advocated by Crouch et al.[12].

In this model the content of a document or query is represented as a weighted set of content terms. The number of distinct content terms used to describe all the documents in the system compose an n dimensional vector space. Each document (and each query) has an n dimensional coordinate in this space. Similar documents should therefore be close together in this space.

The advantages of this system are identified by Crouch et al. as follows:

- An automated search can be limited to those documents close in the vector space to the query.

- The relevance of documents can be ranked according to their distance from the query.

- Relevance feedback (see section A.4.3) can be achieved by moving the query closer to those found relevant and away from those found irrelevant.

Some disadvantages are also listed:

- Some features such as query-document similarity functions are decided only by the system designer
- Terms are assumed to be independent of one another
- Relationships between terms are not expressed

In these respects the reasons the vector space model differs from the MMT. However there is a similarity in that the semantic difference (e.g. number of relationships traversed) between representations (such as textual terms) in the MMT may be used to determine a document's relevance to a query.

A process called *agglomerative hierarchic clustering* can be used to generate clusters of closely related documents, and continue until all the documents are held in one hierarchical cluster structure. This is an alternative approach to the thesaurus paradigm in which terms (or in the case of the MMT, representations) are clustered hierarchically, and the documents indexed on these.

To construct this hierarchy first each of the documents is considered an individual cluster. The two closest clusters are successively merged until one cluster remains. The resulting structure can be represented as a tree with the documents as leaves and the nodes as clusters. An example is shown in figure A.7.

At each node is a figure between 0 and 1 indicating the similarity of the children. The similarity between documents B, C, and D is 0.9 indicating that these three documents cover similar areas. The similarity between that cluster and document A is only 0.7. This figure is useful for controlling the broadness or narrowness of a query.

The produced clusters simplify the matching phase of a query. The search can be performed top-down or bottom-up. During a top-down search the cluster most similar to the query can be traversed until the cluster is more similar to the query than any of the non-document children, or until all the children are documents. The documents in this cluster are those closest to the query, and the query can be broadened by including

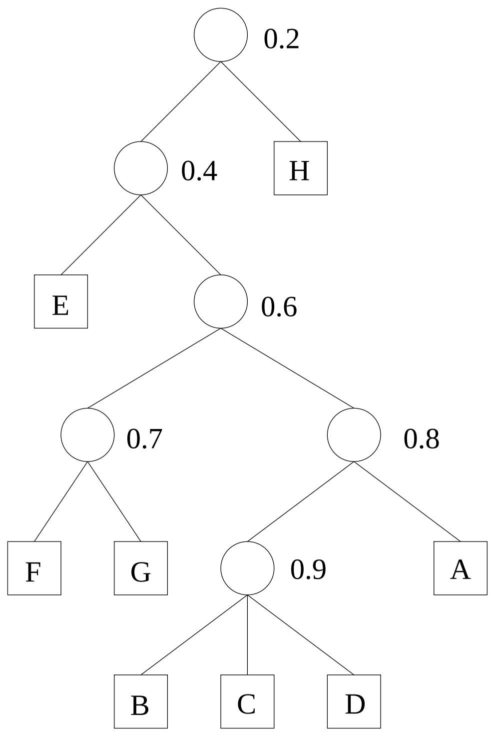

Figure A.7: Sample Document Cluster Hierarchy

parent or child clusters with high similarity values. A similar technique may be useful for finding the most relevant concept in the MMT for a particular piece of media.

At the start of a bottom-up search, on of the clusters at the lowest level of the tree, the one with the nearest centroid to the query, is the starting point for the search. The tree is searched upwards until the current cluster is more similar to the query than the parent cluster.

The structure of the MMT is slightly different so this technique is not directly applicable. However a similar method of similarity could be used as a measure of the specificity of a concept in the MMT — the similarity of the representations of that concept could indicate how broad a concept it is.

A sample implementation of this technique using Hypercard is explained in the paper. The documents are clustered using the vector space model, and placed in a hierarchy using the agglomerative hierarchic clustering method. Retrieval and navigation then proceed as follows:

- The user enters a text query which is converted into term vectors. A list of terms and weights is listed on the screen, though the terms are listed as numbers meaningless to the user. Documents relevant to those terms are displayed as well (though again, only as a list of numbers, not as titles.)

- Next, after changing weights and altering the query to their satisfaction, the user clicks on a "Use Query" button which takes them to a graph representation of the document hierarchy. There is a central node which is the users focus of attention, and the immediately surrounding hierarchy is shown. An overview of the entire hierarchy is also available.

  The initial focus of attention can be a node associated with a query term (the weight of that term in the vector space is high for that node) or the root of the tree. The user can navigate about the hierarchy. Each node and document is annotated with a number indicating how many terms from the original query are relevant to that node or document.

This implementation does have some interesting features which may be useful for MAVIS 2. A graphical representation of the MMT may be useful for navigating through it, and indicating in some way how close each representation or concept is to the query shown may aid the user and stop them unintentionally straying too far from concepts relevant to their query. However concepts should be given a more useful representation on screen than a number, as should documents.

The clustering technique still applies only to text. A similar technique could be applied to any media for which some similarity measure exists, though since these measures are likely to be media and application specific it is unlikely to be suitable for a core method of retrieval and navigation.

A.5.6 Query by Navigation

Another approach to information retrieval and navigation is *query by navigation.* Specifying an information need in the form a of query is one of the key tasks in effective information retrieval, and one that takes experience to be able to perform effectively[46]. The aim of this approach is to introduce the ease of hypermedia browsing to query formulation.

In Golovchinsky's *Queries-R-Links* system, the user can select one or more text terms, usually by clicking on them in a similar manner to hypertext link following [23].

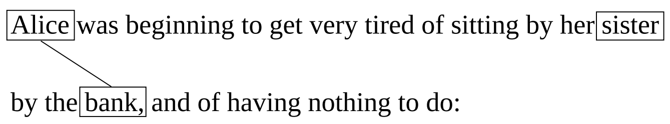

Figure A.8: Query Formulation in Queries-R-Links

Selecting a term adds it to the query, and new set of documents satisfying the query displayed. The user can specify boolean queries by drawing lines between the "and" terms. "Or" is implicit when no line exists. See figure A.8 for an example of this. In the example documents must contain both "Alice" and "bank", OR contain just "sister." A document containing "Alice" is not returned; a document containing just "sister" is, and so is a document containing "Alice" and "bank."

The user can quickly add terms to the query by selecting new terms and connecting them to existing terms if appropriate. A derivative of this system has been used by Golovchinsky to prototype a newspaper archive information system [22].

An alternative approach utilises a *hyperindex* [6, 8]. The hyperindex is also known as the *search space* and allows the user to refine or expand their query by navigating hypertext-style. This search space is similar in form to a thesaurus but is implicit. Thesaurus relations are used in the system.

A small example search space is shown in figure A.9. Unlike a thesaurus, where documents are indexed at the narrowest appropriate term, documents are indexed at *every* appropriate node.

Berger et al. have explained how the behaviour of the searcher can be modelled and estimated by transforming the search space into a transition network [6]. Bosman et al. compare the effectiveness of hyperindices and thesauri for "information disclosure" and find the hyperindex to be at least as useful; However the hyperindex is very text-oriented and can be implemented *on top of* a thesaurus. Applying the idea to the MMT may prove a worthwhile future research topic.

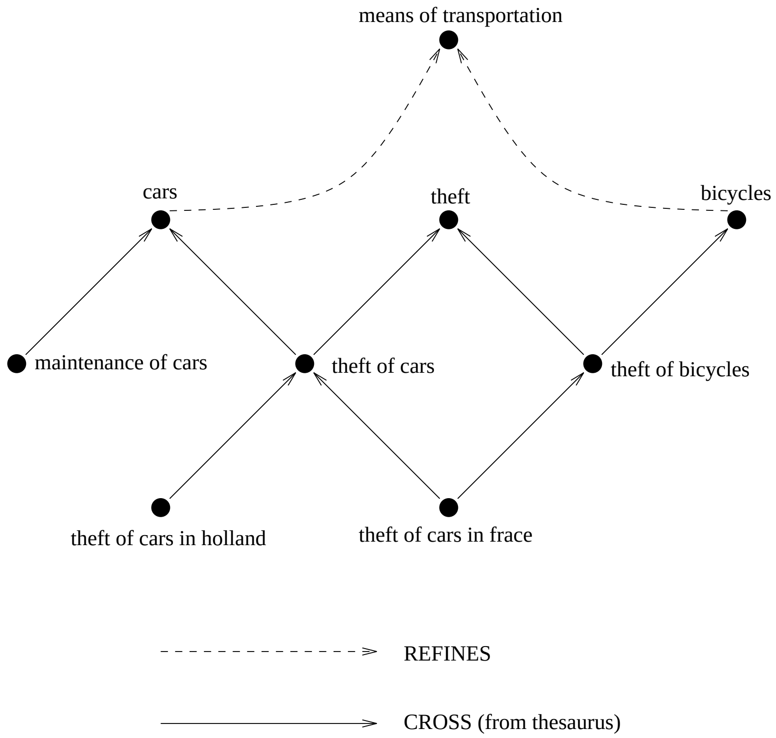

Figure A.9: Query by Navigation Search Space

Bibliography

[1] Jean Aitchison and Alan Gilchrist. *Thesaurus Construction - A Practical Manual*. Aslib, 1972.

[2] Guiseppe Amato, Gianni Mainetto, Pasquale Savino, and Pavel Zezula. Modelling multimedia objects for content-based retrieval. In *Proceedings of the 9th ECRIM Database Group Workshop on Multimedia Database Systems*, Darmstadt, Germany, March 1996.

[3] The art and architecture thesaurus browser. *http://www.ahip.getty.edu/aat_browser*.

[4] M. Beaulieu. Experiments on interfaces to support query expansion. *Journal of Documentation*, 53(1):8–19, January 1997.

[5] Nechemia Daniel Beitner. *Microcosm++: Development of a Loosely Coupled Object Based Architecture for Open Hypermedia Systems*. PhD thesis, University of Southampton, 1995.

[6] F. C. Berger, A. H. M. ter Hofstede, and Th. P. van der Weide. Supporting query by navigation. Technical Report CSI-R9403, University of Nijmegen, Computing Science Institute, Nijmegen, The Netherlands, February 1994.

[7] Colin L. Bird and Peter J. Elliot. Search techniques within a multiple database environment. In John P. Eakins, David J. Harper, and Joeman Jose, editors, *The Challenge of Image Retrieval: Papers presented at a Workshop on Image Retrieval*, pages 1–7, University of Northumbria at Newcastle, Newcastle upon Tyne, United Kingdom, February 1998.

[8] F.J.M. Bosman, R.W.T. Bouwman, and P.D. Bruza. The effectiveness of navigable disclosure systems. In G.A.M. Kempen, editor, *Proceedings of the Informatiewetenschap 1991 conference*, Nijmegen, The Netherlands, 1991.

[9] Artur Caetano and Nuno Guimar aes. A model for content representation of multimedia information. In John P. Eakins, David J. Harper, and Joeman Jose, editors, *The Challenge of Image Retrieval: Papers presented at a Workshop on Image Retrieval*, pages 1–8, University of Northumbria at Newcastle, Newcastle upon Tyne, United Kingdom, February 1998.

[10] J.P. Callan, W. B. Croft, and S. M. Harding. The inquery retrieval system. In *Proceedings of the 3rd International Conference on Database and Expert System Applications*, pages 78–83, 1992.

[11] Peter Clitherow, Doug Riecken, and Michael Muller. Visar: A system for inference and navigation in hypertext. In *ACM Hypertext 89 Proceedings*, pages 294–304, November 1989.

[12] Donald B. Crouch, Carolyn J. Crouch, and Glenn Andreas. The use of cluster hiearachies in hypertext information retrieval. In *ACM Hypertext 89 Proceedings*, pages 225–237, November 1989.

[13] R.M. Crowder, G. Wills, I. Heath, and W. Hall. The application of hypermedia in the factory information environment. In *Proceedings of the IEEE Factory 2000 Conference, Cambridge UK*, pages 411–415, 1997.

[14] Danial Cunliffe, Carl Taylor, and Douglas Tudhope. Query-based navigation in semantically indexed hypermedia. In *ACM Hypertext 97 Proceedings*, pages 87–95, 1997.

[15] Jonathan Dale. A mobile agent architecture to support distributed resource information management. Master's thesis, University of Southampton, May 1996.

[16] S. T. Dumais, G. W. Furnas, and T. K. Landauer. Using latent semantic analysis to improve access to textual information. In *ACM Computer-Human Interaction '88 Proceedings*, pages 281–283, 1988.

[17] Excalibur's visual retrievalware. *http://www.excalibur.be/Gb/products/vrw.htm.*

[18] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, and W. Niblack. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3:231–262, 1994.

[19] A. Fountain, W. Hall, I. Heath, and H. Davis. Microcosm: An open model with dynamic linking. In A. Rizk, N. Streitz, and J. Andre, editors, *Hypertext: Concepts, Systems and Applications*, pages 298–311, 1990.

[20] George W. Furnas and Jeff Zacks. Multitrees: Enriching and reusing hierarchical structure. In *CHI '94 Proceedings*, pages 330–336, Boston, Massachusetts, USA, April 1994.

[21] Brain R. Gaines and Mildred L. G. Shaw. Concept maps as hypermedia components. Technical report, Knowledge Science Institute, University of Calgary, 1995.

[22] Gene Golovchinsky. What the query told the link: The integration of hypertext and information retrieval. In *ACM Hypertext 97 Proceedings*, pages 67–74, 1997.

[23] Gene Golovchinsky and Mark Chignell. Queries-r-links: Graphical markup for text navigation. In *INTERCHI '93 Proceedings*, pages 454–460, 1993.

[24] Willian I. Grosky. Multimedia information systems. Technical report, Wayne State University, 1994.

[25] W. Hall. Ending the tyranny of the button. *IEEE Multimedia*, 1(1):60–68, 1994.

[26] Wendy Hall, Paul Lewis, and Hugh Davis. Multimedia thesaurus and intelligent agent support for content based navigation and retrieval. Technical report, University of Southampton, October 1995. Initial MAVIS 2 proposal to EPSRC.

[27] Wendy Hall, Paul Lewis, and Hugh Davis. Multimedia thesaurus and intelligent agent support for content based navigation and retrieval. Technical report, University of Southampton, February 1996. Updated MAVIS 2 proposal to EPSRC.

[28] Kyoji Hirata, Yoshinoi Hara, Hajime Takano, and Shigehito Kawasaki. Content-oriented integration in hypermedia systems. In *ACM Hypertext 96 Proceedings*, pages 11–21, 1996.

[29] Kyoji Hirata, Sougata Mukherjea, Yusaku Okamura, Wen-Syan Li, and Yoshinori Hara. Object-based navigation: An intuitive navigation style for content-oriented

integration environment. In *ACM Hypertext 97 Proceedings*, pages 75–86, 1997.

[30] R. Jain, S. Murthy, P. Chen, and S. Chatterjee. Similarity measures for image databases. In *SPIE Proceedings, Storage and Retrieval for Image and Video Databases*, pages 58–65, 1995.

[31] Y. Jing and W. B. Croft. An association thesaurus for information retrieval. Umass technical report 94-17, Centre for Intelligent Information Retrieval, UMASS - Amherst, 1994.

[32] K. Spark Jones and R. M. Needham. Automatic term classification and retrieval. *Information Processing and Management*, 4:91–100, 1968.

[33] Susan Jones, Mike Gatford, and Micheline Hancock-Beaulieu. Support strategies for interactive thesaurus navigation. In H. Albrechtsen and A. Oernager, editors, *Knowledge Organization and quality management. Indeks Verlag, Frankfurt*, pages 366–373, 1994.

[34] Dan Joyce. A synopsis of multi-agent systems in mavis 2. Early internal note on intelligent agents in MAVIS 2, April 1997.

[35] B. Azvine K.C. Tsui and M. Plumbley. The roles of neural and evolutionary computing in intelligent software systems. *BT Technology Journal*, 14(4):46–54, October 1996.

[36] Paul H. Lewis, Hugh C. Davis, Mark R. Dobie, and Wendy Hall. Towards multimedia thesaurus support for media-based navigation. In Arnold W. M. Smeulders and Ramesh Jain, editors, *Image Databases and Multimedia Search*, number 8 in Software Engineering and Knowledge Engineering, pages 111–118. World Scientific, 1997.

[37] Paul H. Lewis, Hugh C. Davis, Steve R. Griffiths, Wendy Hall, and Rob J. Wilkins. Media-based navigation with generic links. In *ACM Hypertext 96 Proceedings*, pages 215–223, 1996.

[38] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81–97, 1956.

[39] Institution of Electrical Engineers. *Inspec Thesaurus 1991.* Institution of Electrical Engineers, 1991.

[40] M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra, and T. S. Huang. Supporting similarity queries in mars. In *Proceedings of ACM Multimedia '97*, pages 403–413, Seattle, Washington, USA, November 1997.

[41] S. R. Ranganathan. *Elements of Library Classification.* Association of Assistant Librarians, 1959.

[42] S. E. Robertson. Overview of the okapi projects. *Journal of Documentation*, 53(1):3–7, January 1997.

[43] S. E. Robertson, S. Walker, and M. Beaulieu. Laboratory experiments with okapi: Participation in the trec program. *Journal of Documentation*, 53(1):20–34, January 1997.

[44] G. Salton. *Automatic Information Organisation and Retrieval.* McGraw-Hill Book Company, 1968.

[45] G. Salton and C. Buckley. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 4(41):288–297, 1990.

[46] G. Salton and M.J.McGill. *Introduction to Modern Information Retrieval.* McGraw Hill Computer Science Series, 1983.

[47] J. R. Smith and S. Chang. Visualseek: A fully automated content-based image query system. In *ACM Multimedia '96 Proceedings*, pages 87–98, Boston, MA, November 1996.

[48] S. Smoliar and H. Zhang. Content-based video indexing and retrieval. *IEEE Multimedia*, 1(2):62–72, 1994.

[49] Stephen W. Smoliar, James D. Baker, Takehiro Nakayama, and Lynn Wilcox. Multimedia search: An authoring perspective. In *Image Databases and Multimedia Search: Proceedings of the First International Workshop, IDB-MMS '96*, pages 1–8, Amsterdam, The Netherlands, August 1996.

[50] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing, Analysis and Machine Vision*. Chapman & Hall Computing, 1993.

[51] John F. Sowa. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, 1984.

[52] B. C. Vickery. *Faceted Classification*. Aslib, 1960.

[53] Virage's visual information retrieval tool for informix dbms. *http://www.informix.com/informix/bussol/iusdb/databld/dbtech/sheets/virage.htm*.

[54] Visualseek content-based search engine. *http://www.ctr.columbia.edu/ jrsmith/VisualSEEk/VisualSEEk.html*.

[55] S. Walker and R. De Vere. Relevance feedback and query expansion. In *Improving subject retrieval in online catalogues*, number 2 in British Library research paper. British Library. Research and Development Department, 1990.