A Flexible Architecture for Content and Concept Based Multimedia Information Exploration

Mark Dobie, Robert Tansley, Dan Joyce, Mark Weal, Paul Lewis, Wendy Hall
Multimedia Research Group
Department of Electronics and Computer Science
University of Southampton, Southampton, SO17 1BJ, Great Britain
Email: mrd@ecs.soton.ac.uk

October 30, 1998

Abstract

Traditional hypermedia systems can be extended to allow content based matching to give more flexibility for user navigation, but this approach is still limited by the capabilities of multimedia matching technology. The addition of a multimedia thesaurus can overcome some of these limitations by allowing multimedia representations of concepts to act like synonyms in the query process. In addition, relationships between concepts allow navigation within the context of a semantic scope. The use of agents that independently examine the information in the system can also provide alternative methods for query evaluation. This paper presents a flexible architecture that supports such a system and describes initial work on implementation.

KEYWORDS: Content based navigation, multimedia architecture, thesaurus.

1 Introduction

Traditional hypertext systems allow textual information to be arbitrarily linked so that users can navigate between related parts of the information in the system. Many of these systems can use multimedia information such as images, sounds and video clips that can also be linked. In such hypermedia systems the links are usually created using specific locations in particular documents.

Some systems, such as Microcosm [6], also allow links to be created by specifying the text that forms one anchor of the link. These links are created once and can be followed from any location where the text occurs. To achieve this, the system has to examine documents currently being viewed by the user and look for matches between text in the documents and text that forms anchors of links within the system. Where matches occur the system can highlight the text as a source anchor for a link.

This is a form of content based navigation. Navigational links are dynamically created by matching the content of currently viewed documents with the content of previously created link anchors. For textual documents and links this is relatively easy to implement since comparing text strings is often straightforward. Content based navigation for multimedia documents and links involves comparing selections of images, video and sound with each other.

This is still an unsolved problem in the general case and implementations are limited in their capabilities. The remainder of this paper begins by describing some of the reasons for these limitations and suggests that a multimedia thesaurus can improve the flexibility of a multimedia information system. It continues to describe some related work in the areas of content based retrieval and thesaurus supported navigation.

A novel architecture for a multimedia information system is proposed together with the design for MAVIS 2, the research system currently being developed by the Multimedia Research Group at the University of Southampton. MAVIS stands for Multimedia Architecture for Video, Image and Sound and this is the second project by that name.

The aim of the MAVIS 2 project is to explore the use of a multimedia thesaurus and intelligent agents to support navigation by content and semantic relationships between concepts in multimedia information systems. MAVIS 2 is currently being implemented and some discussion of this process is given with prototypes of implementation so far.

1.1 The Multimedia Matching Problem

The problem of matching multimedia selections is addressed by content based retrieval systems [4, 14, 9]. The aim of these systems is to find objects that are similar to an example object indicated by the user. There are several aspects to the multimedia matching problem that make it a challenge that remains unresolved.

1.1.1 Properties

For each medium there are different properties that can be measured that represent different ways in which two objects can be considered similar. For example, images can be compared on the basis of colour, shape or texture and audio segments can be compared by frequency, tempo or rhythm. When a user asks for a similar object the appropriate types of similarity have to be indicated.

1.1.2 Signatures

Each property can be represented by different signatures or feature vectors and there are different algorithms for calculating and comparing them that may be appropriate in different situations.

1.1.3 Similarity

It is often the case that the properties that a computer can use to compare objects do not correspond to the properties that humans use to compare objects. Humans can apply world knowledge to recognise objects in scenes and sounds in audio that requires a level of media understanding that computers cannot achieve yet. Content based retrieval systems have to combine results from existing measures and attempt to match objects.

Despite these problems there are many content based retrieval systems, existing as both stand alone systems and incorporated into database products.

The techniques of content based retrieval can be used in a hypermedia system to provide multimedia content based navigation [11]. This allows previously created links to be made available to the user where the objects that the user is viewing are similar to the objects from which the original links were made, subject to the constraints of the underlying matching methods. This allows many links to be made available to the user without the author having to create them explicitly.

1.2 Thesaurus Support for Navigation

Content based navigation provides a useful extension to a hypermedia system, but it still requires that there is some sort of similarity between link anchors and selections within users' documents. There are many situations where a given link might be appropriate but the matching methods cannot see the similarity between a selection in a document and a link anchor. Different views of the same object or different examples of the same type of object may appear dissimilar.

For example, a top view and a side view of the same table look very different. An office chair and an easy chair are not similar in shape, but they are both chairs. Similar examples exist for other media, such as words spoken in a different voice, alternative versions of the same song, the same object moving differently in a video clip.

Some of these relationships can be represented by using a thesaurus. For a given term a traditional thesaurus gives synonyms (words with similar meanings) and semantically related terms (usually broader and narrower terms).

The thesaurus idea can be extended to include multimedia information about each term. It can include different multimedia representations with each term that can be used to illustrate different views, types or aspects of a concept. The particular representations that are appropriate will depend on the application for which the thesaurus is designed.

When such a multimedia thesaurus (MMT) is combined with multimedia matching technology it is possible to offer more flexible facilities for content based navigation and retrieval [10]. For example, if a query matches with one representation of a concept then information associated with other representations can be offered to the user.

Concepts in the thesaurus are related semantically, so it is also possible to express the scope of a search as broader and narrower term relationships. Another possibility is an automatic synonym search, where a match with a thesaurus entry can trigger parallel searches using synonym multimedia representations from the thesaurus as queries for the searches. The user can match against representations associated with thesaurus concepts and then use the structure of the thesaurus to explore the semantic relationships between concepts.

The combination of hypermedia links, content based matching and a multimedia thesaurus offer numerous possibilities for the user looking for information. One challenge presented by this work is devising a user interface that allows the user to express clearly how they wish to use the system while keeping simple operations simple to perform.

2 Related Work

Multimedia information retrieval is a widely researched area and there are many commercially available database systems with this functionality. This section briefly describes some of the systems that are available.

QBIC

The Query by Image Content (*QBIC*) system developed by IBM [4, 15] allows images from a collection to be retrieved via visual cues such as colour distribution or texture. This searching can be combined with other cues such as keywords. Though it is purely a retrieval system and does not include hypermedia functionality, it does face the same matching problems described earlier. QBIC has been integrated into IBM's DB2 database in the form of multimedia extenders that allow a content based multimedia match to form part of a normal database query.

Informix Visual Information Retrieval DataBlade Module

The Informix Software DBMS can be extended with a Visual Information Retrieval system, allowing images to be retrieved via visual cues. Keywords can be used to aid the search. These systems can retrieve images by similarity measure, but do not take into account any semantic information other than keywords. Used as a supplement to hypermedia functionality however they become a valuable navigation tool.

The MAVIS project

The original MAVIS project [11] extended Microcosm's generic link capability [6] to allow source anchors to exist in media other than text. The user can retrieve the source anchors of links via media-specific clues (such as colour distribution or texture for images). The source anchors in the system are then searched and a ranked list of matching source anchors is presented, allowing the user to follow each link.

Semantically Linked Hypermedia

Use of semantic information in hypermedia/information retrieval queries is an ongoing research topic. Cunliffe et. al. have proposed a semantic hypermedia architecture [3]. Binary relationships between index terms or media item identifiers are stored in a Binary Relational Store. The user can query this by specifying a piece of media held in the store and optionally a relationship. Relationships pertaining to that piece of media and any specified relationship will then be presented to the user.

Content-Oriented Integration

The Content-Oriented Integration work by the C & C Laboratories, NEC USA Inc. [7, 8] allows the user to search for media via media-based clues, as well as allowing conceptual-style navigation. Media items are connected to conceptual representations held in the system. These conceptual representations are linked to each other allowing navigation on a conceptual level as well as a media level. It also allows *object-based navigation* whereby a number of objects and their relationships can be used in a query. Conceptual navigation is much like MMT navigation in MAVIS 2.

Link Clustering

Link Clustering is a technique with similar goals to the MMT in MAVIS 2. A link cluster has several source anchors and several destinations. If any of the source anchors are reached all of the destinations of the cluster are presented to the user. In addition, arbitrary relationships are allowed between clusters that are available for navigation between clusters after a cluster has been reached. This technique is currently being researched by Crowder et. al. [2].

The Okapi Project

Retrieval using text-based thesauri has been researched for decades, though widespread use of electronic thesauri for this purpose developed more recently. Thesauri are typically used either as a means of indexing documents according to the preferred form of a particular term, or used for synonym substitution in a flat text search.

The Okapi project researched at City University London [12] involves the testing of various text-based information retrieval techniques including the use of thesauri. Beaulieu [1] concluded that thesaurus use, both explicit and implicit, was beneficial to the retrieval process.

3 MAVIS 2 Design

MAVIS 2 uses a four-layered data architecture to describe the objects that for the relationships within the system. This is shown in figure 1 and described in the following sections.

3.1 Raw Media Layer

This consists of representations of all the raw media objects that are usable within the system. A Raw Media object contains a reference to a media file such as a web page or an image file and some information about the type of the object. Raw Media objects are referred to by higher level selection objects.

3.2 Selection Layer

The selection layer contains selection objects that describe part of a Raw Media object. For example, a selection may refer to an area of an image, an extent in a piece of text or a time segment in a piece of music. Several selections can refer to different parts of the same Raw Media object, removing the need to duplicate large media data.

In order to compare two selections, signatures (feature vectors) are derived from the portion of raw media to which the selections refer and the signatures are compared with each other. Signatures are not stored explicitly at this layer but are calculated and cached with an association to the selection from which they were derived. A selection may have many signatures depending on the feature extraction algorithms that are available and any two selections can be matched by comparing any or all of the types of signatures that they have in common.

3.3 Selection Expression Layer

This layer consists of selection expressions that describe combinations of selections and information about which properties of each selection are relevant. Simple combinations of selections may be constructed by means of a GUI, for example *the colour of selection S1* and *the shape of selection S2*. By default, all properties of an unqualified selection are considered relevant.

In MAVIS 2 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 attached to it for application-specific purposes.

3.4 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 one or both of of two types

Lead-in

The selection expression will by default be used for matching with a view to reaching the concept. Also it will be offered to the user when they ask for alternative representations (like synonyms in a text thesaurus) of the concept.

These are analogous to the textual terms in a controlled language thesaurus, whereby equivalent terms “lead in” to the preferred term for that concept. However, in the MMT the concept itself is an abstract representation, thus the preferred form is a lead-in itself. This preferred form is flagged as the *preferred representation*, and typically associates a selection expression consisting of a single textual term with the concept. This term is then displayed when the concept needs a concrete representation, for example in a concept browser.

Relevant document

The selection expression will by default be offered to the user when they ask “tell me about this concept.” These are analogous to documents indexed by a traditional thesaurus.

A selection expression can be both a lead-in and a relevant document, in which case it is possible to match with it to find a concept and to use it to illustrate a concept. An example might be a photograph of someone associated with a concept that represents that person.

3.5 Conceptual Layer

This consists of abstract representations of concepts and no media or keywords are held in this layer.

Relationships between these concepts are represented by *Concept/Concept relationships* which 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.

Though this layer by default will use classic thesaurus-style relationships (broader/narrower, related) a richer set of semantic relationships can be expressed by defining new sets of relationships. An example might be spatial relationships to describe the relative locations of objects with respect to each other.

4 MAVIS 2 Queries

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

The query is processed and the results are presented to the user. The user can interact with the results and find out why each result is present and also reorder the results based on various criteria. The user can then either follow links, view other selection expressions or view and browse concepts in the MMT.

The three main parts to a MAVIS 2 query are described in the following sections.

4.1 A Query Selection Expression

The user specifies a *query selection expression* that specifies a combination of selections and properties against which other selection expressions that fall into the scope of the query will be matched. An initial tool for building selection expressions allows the user to specify a set of selections and properties that they wish results to be similar to and a set of selections and properties that they wish results to be different from. In other words, the user can effectively say "I'm interested in objects like this, but not like this."

After the query has been processed this query selection expression will be modifiable, allowing the user to refine their query, perhaps augmenting it with selection expressions that were results from a previous query. 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 added to the set of selections in the query expression that the user is not interested in matching.

4.2 A Query Scope

Usually the user will have some idea of what they want from the query. It is the scope of the query that governs which selection expressions in the system are compared with the query selection expression. 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 that have been made 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.

4.3 MMT/Agent 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.

The use of the MMT to expand queries may be explicit or implicit. Both techniques have been shown to be useful in text-based retrieval scenarios [1].

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

The user can also control how intelligent agents are utilised in the matching process. These agents interact with the query process by receiving signature vectors related to various terms in a selection expression and processing them to efficiently find related signature vectors. Using appropriate media feature extraction techniques, the similarity of media objects can be determined by standard statistical techniques. The agent's role will then be:

- To discover novel associations and build schema relating selections in the system (for example, examples of similarly coloured tables) with each other using single or multiple feature extraction techniques.
- To efficiently retrieve selections which are deemed to be of the same classification given examples of other objects (for example, given a colour histogram feature vector of a red table, efficiently find other objects with similar colour histograms).

It is also possible for an agent to make use of the explicit associations between selection expressions and concepts in the MMT to provide a direct mapping from feature vectors to concepts that can be more efficient than the exhaustive nearest neighbour comparisons that are otherwise required.

4.4 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 and two selections are compared by comparing signatures (feature vectors) derived from them.

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 reflects the relative strength of a match rather than an absolute number calculated by an algorithm in an arbitrary way.

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.

The results of a query can be presented to the user as a ranked list. An abbreviated representation of the matching selection expression may 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 frequency distribution of match qualities of each property will be available.

From this list the user can find similar selection expressions, view the destinations of hypermedia links or explore concepts which are appropriate to the query.

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 [5, 13].

5 Prototype Implementation

As a research system, one of the goals for the implementation of MAVIS 2 is do develop a system that is flexible enough to allow different parts of the system to be implemented in different ways and to accommodate the varying functionality that these implementations may provide.

5.1 Messaging Architecture

The basis of the implementation is a flexible messaging architecture and storage system that allows any part of the system access to the information that it requires. All operations are initiated by sending messages which allows for easy distribution of the whole system across a network.

The design is based around a set of communicating processes that operate independently from each other. The functionality of each process is described by a group of four values that describe the messages it can receive, the inputs it can process, the outputs it can produce and the name of the process. When each process starts it registers as many of these functionality specifications as required (to fully describe its capabilities) with a broker process, along with the address of the process.

In order to perform a given task a message is assembled and sent to the broker process. The broker process matches the requirements of the message to the capabilities of available processes and forwards the messages appropriately. Any of the four parts of a functionality specification may contain wildcards for very flexible matching.

Referring to processes by their functionality allows more than one process to be available for performing a given task. This allows alternative matching methods, such as a computational match and an intelligent agent, to be used in parallel and completely *transparently* to the sender of the message.

If a messages is intended for just one recipient (for example, a message to store an object in the system) then the name of the process can specified explicitly. If all the processes are uniquely named then only one process will match and receive the message.

MAVIS messages are encoded using XML and transmitted between processes using HTTP. This gives an automatic remote addressing scheme in the form of URLs and allows integration with other systems that provide an HTTP interface. Each process is implemented in Java and contains a servlet that allows it to act as a web client and server in order to send and receive MAVIS messages.

5.2 Storage System

A simple storage system has been implemented that allows MAVIS data objects to be stored and retrieved by ID and by content. Both the ID and content parameter values can have wildcards in them, allowing for subsets of the stored objects to be retrieved based on ID, object type and references to other objects.

In this way it is relatively easy to form requests such as "all signatures that were derived from this selection" or "all selections of this type".

In the future this storage process may be re-implemented to take advantage of the indexing and retrieval facilities offered by a dedicated database system, but this can occur without changing the messaging interface that is presented to the rest of the MAVIS 2 system.

5.3 Processes

A set of eleven processes has been implemented within the architecture described to provide the evaluation of simple queries. The broker and storage processes provide the basic facilities that the other processes require.

A simple viewer that allows selection from web pages is used to import selections into the system and initiate queries. In the initial version selections are text and queries consist of a single selection. A query evaluator receives a query and assembles a list of selections in the system that fall within the scope of the query.

These are passed to a selection matcher that derives (or looks up previously calculated) signatures from the selections to be compared and passes them on to an appropriate signature matcher. When a signature needs to be calculated a message is sent to a signature module process. It is possible that different types of signature may require different methods in order to compare them.

The signature matcher produces lists of distances in feature space for each type of signature that are passed to a results generator, which collates the distances into lists based on the object from which they were derived. A results viewer receives these lists and orders the results before presenting them to the user. The user can select individual results and use them as the basis for further queries.

The multimedia thesaurus is a separate process and when it is active (if the user sets the appropriate scope settings in the viewer) it receives the same query message as the query evaluator and transparently provides support for the query. The initial version can return concepts as results and it can initiate its own parallel searches on synonyms of the query to further augment the overall list of results.

There is also a concept browser and editor that allows the concept network used by the MMT to be viewed, navigated and modified.

6 Conclusions and Further Work

In this paper we have presented an architecture that allows a multimedia thesaurus and agents to be combined with a content based hypermedia system, with the aim of overcoming some of the limitations of multimedia matching and improving the flexibility of content based navigation.

An initial prototype has been implemented that successfully demonstrates the flexibility of the messaging and storage systems by evaluating simple text based queries with optional and transparent support from the multimedia thesaurus using a set of independent processes distributed across a network of machines.

Separate research has been conducted on methods for flexible selection from images and image based signatures to use for comparison of image selections and this work can be integrated into the framework that has been established. Other research in the group is concentrating on selection methods and signatures for audio and video information.

In addition, separate research on suitable intelligent agents will be added in a similar way to the MMT and further work is being carried out that will allow the facilities of MAVIS 2 to be available from a standard web browsing environment.

References

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

[2] 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.

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

[4] 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.

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

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

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

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

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

[15] H. Treat, E. Ort, M. Vo, J.S. Jang, L. Hall, F. Tung, and D. Petkovic. Searching images using ultimedia manager. In *SPIE Proceedings, Storage and Retrieval for Image and Video Databases*, volume 3, pages 204–213, 1995.

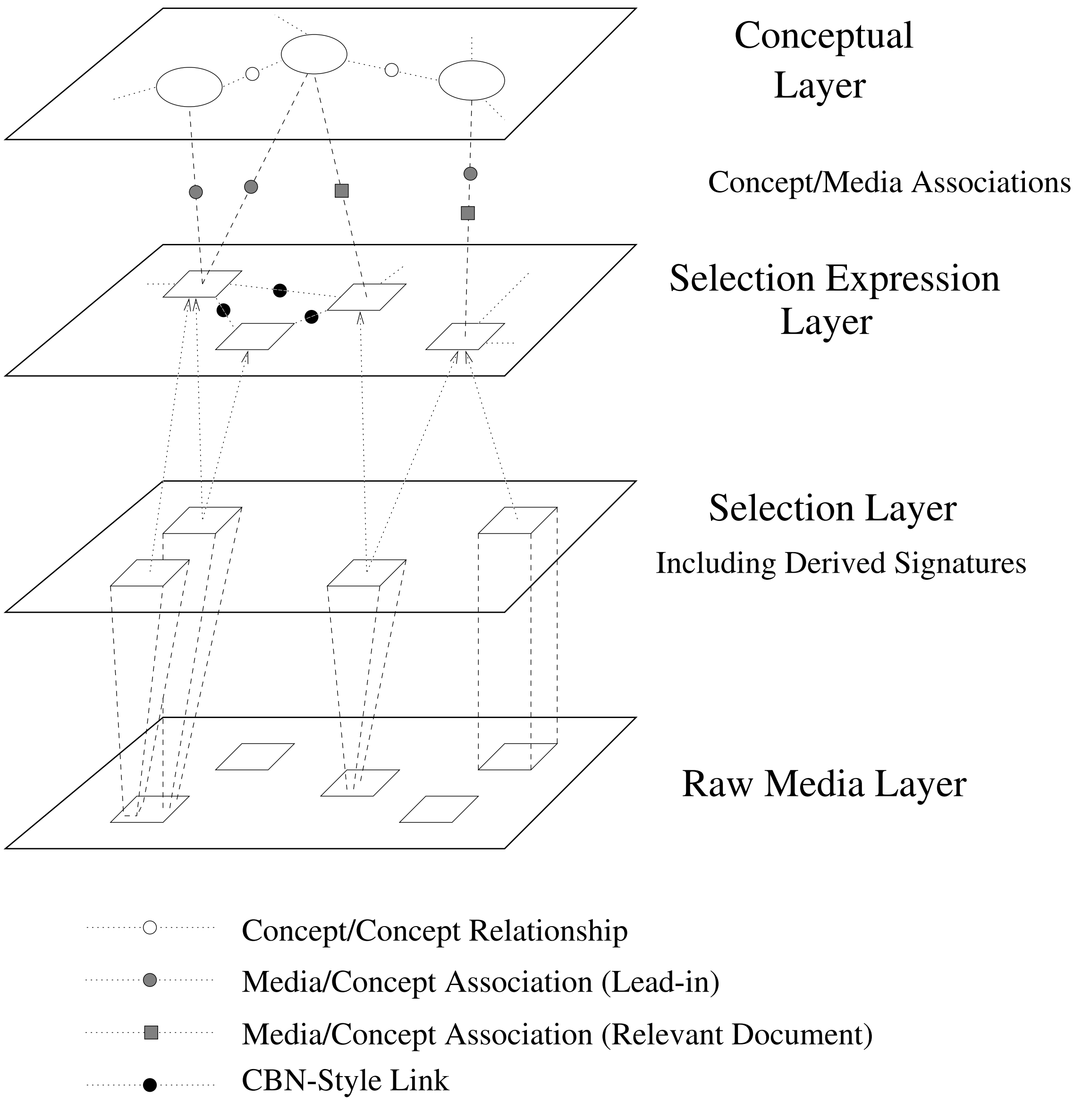

Figure 1: Four-Layer MAVIS 2 Data Architecture