State Of The Art

In the context of this thesis' motivation and contemporary research work, this chapter evaluates the specified requirements for intelligent multimedia generation with respect to existing system implementations. While the IMMPS Standard Reference Model is seen to support the requirements, we demonstrate that existing systems do not. We conclude with a taxonomic summary of these failings.

The motivation of this thesis is to respond to the hypothesized future of pervasive access to digital content with intelligent multimedia presentation, enabled through an automated and adaptive system. As a result of defining and analysing two typical scenarios in this hypothesized future, a set of fundamental requirements for systems realizing a multimedia generation process was determined (section 1.4):

- The retrieval of data from different sources
- The processing of data of heterogeneous type and format
- The incorporation of contextual adaptation throughout the process
- The dynamic integration of external knowledge
- The presentation of multimedia on the basis of expressed concepts and their relationships

Having introduced how the evolution of computer systems from syntactically-based operation to semantically-based operation can be described as a movement from the "second generation" to the "third generation" (section 2.4), we have described the problems encountered by the second generation of systems for multimedia presentation generation. In describing how the "third generation" Web is being realized, we indicated the potential that this approach may have in resolving those problems.

The basis for such "third generation" systems for multimedia presentation generation is the utilization of knowledge in the generation process. The Standard Reference Model for Intelligent Multimedia Presentation Systems (SRM-IMMPS) has been described (section 2.5), as it is intended to be a common reference architecture for such "third generation" systems.

We will consider if the given requirements are met by the SRM and compare this to existing multimedia system implementations that can be considered examples of such "third generation" systems (i.e. that they incorporate a knowledge-based aspect to the multimedia generation process). In this chapter we consider the following systems (in order of being mentioned in the text):

- MIPS [Wilson,1997]
- Artiste [Allen,2002]
- MADEUS [Jourdan,1997]
- Otto et al [Otto,1998]
- Cuypers [van Ossenbruggen,2001]
- MAVIS 2 [Tansley,2000]
- MediaNet [Benitez,2000]
- Shih's IMMPS [Shih,1996;Shih,1997]

These systems represent all those found in the literature which were developed to enable the automated (to some extent) generation of multimedia presentations in a domain generic way and which both refer to the SRM-IMMPS in their work and utilize (again, to some extent) semantics in their approach (as is necessary if automatable multimedia generation is possible, as noted in section 2.1.4). It is interesting to note that the literature dates from 1997 to 2002, it is not that there has been no further research in this area but there have not been any significant new results about the systems themselves. Much work was not continued, while Artiste was superceded by SCULPTEUR (which continued the same approach described here) and Cuypers continues to be used as a basis in current projects without significant changes in its process. Where later literature has introduced new relevant results, it is cited in the text.

3.1 Retrieval of data from different sources

Given the hypothesized pervasive accessibility to digital content regardless of content location, source, network or delivery the desired multimedia presentation system should be able to retrieve data from any potential source. Furthermore given that related content is likely to exist in different locations and sources, on a range of networks and delivery mechanisms, the system must support data acquisition from multiple, heterogeneous sources.

The SRM supports the use of external content sources. As well as proposing "presentation goals and commands" as the system input, the model shows an additional input source, labelled "application", which is explained as indicating external data or knowledge from an application. In the Content Layer it is suggested that a Content Selection component communicates with the Application Expert. The model specifies as part of the tasks of the Application Expert to be interfacing with external applications and the provision of data to the system. Hence the SRM proposes that the logic required by the multimedia system to be able to retrieve data from potential heterogeneous sources should be based upon declarative knowledge modelling in the system knowledge base.

The MIPS system architecture includes a knowledge-based query mapper which expands a single query into a set of concepts and their attributes. When these attributes are themselves concepts, these are further expanded in a recursive manner. Otherwise the attributes are linked directly to "ground types", i.e. media items that represent the original concept. These direct links can be made into any content source able to be referenced, however the retrieval mechanism must be implicit in the link, e.g. a HTTP type URL. Therefore it must be considered that the system is not easily extendible to support other retrieval mechanisms which can only be implemented within the system logic and identified by a link type. That content is directly linked builds into the system a lack of flexibility in content selection (e.g. the filtering of media selection relating to a given concept on the basis of additional factors) and in automated link generation (the concept-media links are manually set and altered). Indeed, the entire query mapping tool is internally authored to support the domain being used in the system, which is a non-scalable approach and not flexible in the support of other domains.

The Artiste project had the aim to develop a cross-collection search system for art galleries and museums. In order to support media retrieval from multiple heterogeneous sources, it takes the approach of using internal mappings across the different database schemas. A single query is then mapped to produce multiple equivalent queries to be applied on the multiple sources. In order to support queries more complex than low-level textual retrieval, the query items are modelled in a knowledge representation. Media queries are modelled with the concepts represented by the media, properties such as colour or shape, and attributes such as size. Such internal mappings are effective for enabling content retrieval from all sources which are using schemas supported by the mapping tool. This approach lacks scalability however as potential content sources increase (unless some level of schema uniformity is certain) and the dynamic incorporation of external mapping information is not supported. Content retrieval is also clearly restricted to applications with the required interface (standardized access to a database model).

3.1.1 Summary

To summarize, we note that the majority of the systems are built to accept a single input which contains all that is necessary to produce the final presentation. This is to exclude other data that is distinct from the input data, such as context (for adaptation) or rules (for presentation) which may be retrieved by the system after the initial input – these cases are mentioned in the later sections.

SRM-IMMPS	Foreseen in the Application Expert
Madeus	X
Cuypers	X
MAVIS 2	X
MediaNet	X
Artiste	RDF query mapping to SQL
Shih’s IMMPS	X
MIPS	Internal query mapping
Otto et al	X

3.2 Processing heterogeneous content

Digital content, given the wide range of sources from which the content will be made available and the wide range of content types which shall be offered, will likely be expressed by different data formats and structures. While XML could play the role of a single standardized data structure for representing or pointing to digital content, different vocabularies will exist for different content domains and media representations or descriptions.

The SRM’s definition of the Application Expert includes the knowledge representation for “data exchange and conversion” as well as “understanding of and reasoning on data”. This explicit mention of both data conversion and data reasoning indicates the models presupposition that not only content of different data formats will be handled, but also content that exhibits a knowledge structure as opposed to a data structure (i.e. to state this in third generation Web terms, a semantic model such as RDF rather than a structural model such as XML).

The MADEUS system architecture, for example, uses the XML data model for the representation of an abstract multimedia presentation. The system supports a single XML input, which is transformed using a XSLT stylesheet to form the abstract multimedia presentation from which a final format multimedia presentation can be realized. By using different XSLT stylesheets, different XML formats can be handled by the multimedia generation process. The simple linear transformation process of the system is illustrated below (Fig 3.1).

Figure 3.1 The MADEUS generation process

This approach, while typical in the Web context and able to benefit from the wide usage of XML and tool support (e.g. XSLT), also requires the manual association between XML and XSLT resources and excludes any dynamic selection of transformational rules according to the format of dynamically acquired input documents. This fails to match well with the run-time retrieval of content from multiple, heterogeneous sources.

In the Artiste project, there is a single query, which is mapped into the different supported target vocabularies. The query modelling was realized using the RDF standard, so the mapping could also support queries in the RDF model although the project examples focus on mapping to relational databases using SQL as the final query language. As the mapping is at the semantic level (it operates on the equivalence of vocabulary terms) the data structure of the queried databases must be homogeneous. This mapping is focused on the semantic classes, rather than also supporting the equation of semantic instances, so that it is still necessary for the different sources to share an instance vocabulary. In other words, the use of a semantic model in the projects' approach is not taken further than complementing relational database structures.The Sculpteur project[27], the follow-up to Artiste, implemented a multimedia retrieval system with Semantic Web functionality. It used a RDF(S) based ontology for associating concepts to digital representations of museum collections and reasoning about those collections which combined low-level multimedia retrieval with high-level reasoning, and hence supported Semantic Web interoperability with its ontological approach. However, the core approach does not differ significantly from that of Artiste, as the focus of the project was to extend previous results into large museum collections of both 2D and 3D objects [Addis,2005].

3.2.1 Summary

Some of the systems considered here did choose to use flexible formats such as XML or RDF where mappings can be provided to overcome the problem of heterogeneity. The remaining systems are fixed to single, often proprietary formats which restricts their ability to handle any other formats without specific pre-transformation making the dynamic incorporation of heterogeneous content impossible.

[27] http://www.sculpteurweb.org

SRM-IMMPS	Foreseen in the Application Expert
Madeus	Use of XSLT so limited to structured data
Cuypers	X
MAVIS 2	X
MediaNet	X
Artiste	Limited to relational database schema
Shih's IMMPS	X
MIPS	X
Otto et al	X

3.3 Incorporation of contextual adaptation throughout the process

In processing heterogeneous content to form a coherent multimedia presentation for a wide range of potential contexts, intelligent adaptation is a fundamental requirement of a multimedia presentation system. To ensure the maximum adaptive flexibility there should be a contextual representation made available at run-time to every stage of the multimedia generation process. The process should then, in turn, be able to interpret the contextual representation at every stage in order to be able to introduce contextual adaptation at the appropriate level of the multimedia presentation modelling.

In the SRM two experts in the knowledge server are dedicated to contextual knowledge: the User Expert and the Device Expert. This reflects the particular significance of these two factors for the production of suitable multimedia content. In multimedia system implementations there is often explicit contextual adaptation to either or both of these factors. However the SRM also allows for contextual knowledge beyond that of these factors, including in the Context Expert the intention to resolve "context dependent references" within a stated domain. This much more generic requirement of a knowledge expert indicates the intended generic nature of its implementation, i.e. that just as context can be a reference to a wide range of possible domains – not just that of the user or the device – the knowledge support for contextual adaptation should also be able to support this wide range of domains.

In the MADEUS research the importance of contextual adaptability is recognised and responded to [Roisin,2000]. The MADEUS system extends its syntactic (XML) processing with adaptation support by supporting the inclusion in the XSLT stylesheets of system-understood contextual parameters, which are matched with execution parameters to select the most appropriate XML transformation. The proposed architecture (Fig 3.2) indicates that adaptation parameters could act in different places in the process – at the transformation sheet level, at the template level or as an input to the formatting process.

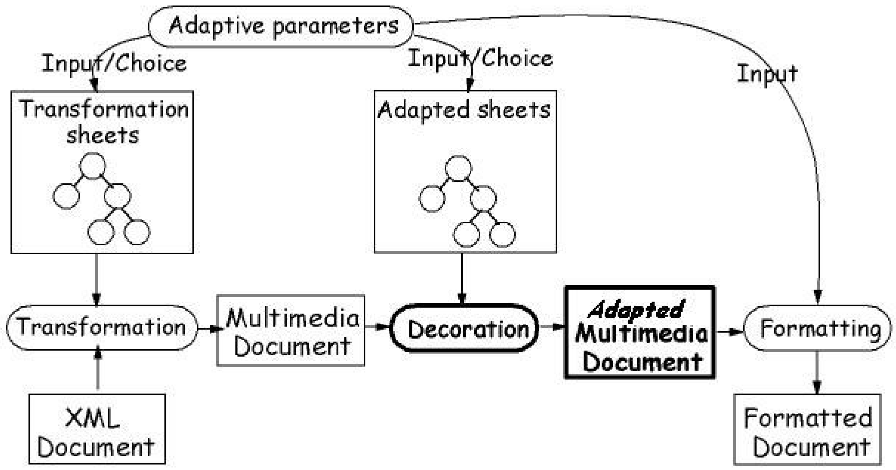

Figure 3.2 MADEUS proposal for adaptive presentation

Additionally an intermediate transformation ("decoration" – in bold in the diagram) could handle syntax-independent adaptations. The representation of the contextual parameters could be system-specified, while in research papers the RDF format CC/PP is mentioned [W3C,2004b], however this format would restrict contextual adaptation to device capabilities.

An approach to implementing this architecture is also described [Bes,2001]. It consists of an 'Analyser' which selects style sheets and constraints sets on the basis of the input context parameters, and a 'Transformer' and 'Optimiser' which applies the style sheet and constraints on the initial XML document to produce the final formatted document.

In this approach the contextual selection is based on knowledge explicit within the processing tools meaning that support for new or altered contextual knowledge can only be realized by altering or producing new XSLT stylesheets. Given that there can only be a reasonably finite number of XSLT stylesheets available to the system to match a much greater range of potential contexts, there is also a question of flexibility in the parameter matching. Tighter matching rules will exclude many contexts but looser matching rules may not produce well-adapted results.

Otto et al have proposed an IMMPS which interprets the act of intelligent multimedia presentation generation as "a projection from the information model to the presentation model with regard to the context model".

The information model defines the concepts in the content domain and their relationship to individual media objects. The context model is the representation of all influencing factors on the presentation generation. The authors mention:

- the characteristics of the information space e.g. objects of the same class are presented equally.
- the characteristics of the media objects e.g. the media type for display
- the characteristics of the user
- the characteristics of the target machine
- the presentation context.

The generation process itself is presented as a three stage process: content selection (on the basis of the users' formulated query), media selection and media realization. The use of the context model is proposed only for the intermediate stage of media selection. The content selection of the first stage is refined over a number of "media set reduction steps", in which the context model is processed by system rules. These rules are simple and easy to check, and are authored in a system-specific interface, using a limited number of test operations (e.g. less than, equals to) on the available attributes from the system model. This approach is then system specific, not interoperable with external knowledge and the contextual adaptation is focused only on the filtering of pre-selected content, and does not support adaptation in terms of altering the initial selection or the modification of selected content.

The Cuypers work incorporates support for contextual models as part of the input to the Cuypers multimedia generation system. The research system explores a set of abstractions, both on the document and presentation level, geared towards interactive, time-based and media-centric presentations. It makes use of extensive transformation rules implemented in Prolog and backtracking support between the transformation steps.

The proposed implementation is a five step transformation from an initial abstract (XML-based) data structure to a final multimedia document, relying on constraint-solving techniques to ensure a result that meets the authoring requirements as closely as possible. This developed out of earlier findings that "for most applications, the conceptual gap between an abstract, presentation-independent document structure and a fully-fledged, final-form multimedia presentation is too big to be specified by a single, direct transformation" [Rutledge,2000]. This five step process is shown below (Fig 3.3).

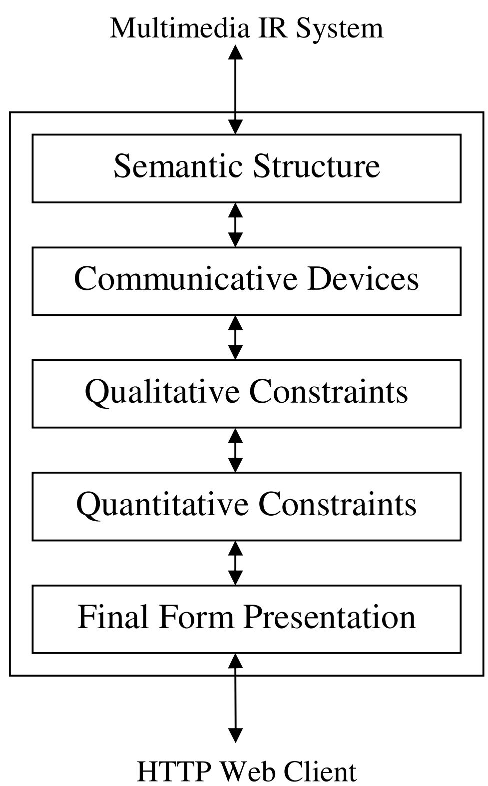

Figure 3.3 The Cuypers generation engine

While the principle input is a “rhetorical structure” [Mann,1987] represented in a system-specific XML format, generated from a single, semi-structured multimedia database, the developers acknowledge the need for additional semantic modelling to the specific context of an individual user [van Ossenbruggen,2002]. Explicitly mentioned as system inputs are models for the knowledge about the user and the device. These models contribute to the formation of the abstract representation of the multimedia presentation by being passed through the linear process of the system, yet as the research work itself acknowledges, the knowledge representation and handling is currently implemented “using ad-hoc encoding representation techniques” and “remains implicit and hidden in the (procedural) generation software”.

3.3.1 Summary

Adaptation is an important requirement for a multimedia presentation system if it is to support delivery to different types of user, device or situation. Adaptation in a system is dependant upon two factors: the availability of context information that can be interpreted and the flexibility of the internal data structures to be altered according to the interpreted context. There is no common agreement on the format of context information. Those systems which have attempted to handle context has chosen their own formats and implemented adaptation in a proprietary and inflexible manner.

SRM-IMMPS	User, Device and Context Expert
Madeus	XSLT for additional adaptation
Cuypers	Additional input handled internally
MAVIS 2	X
MediaNet	X
Artiste	X
Shih’s IMMPS	X
MIPS	X
Otto et al	Internal proprietary context and rules

3.4 Dynamic integration of external knowledge

If the proposed multimedia generation system is to be “intelligent”, it should be able to not only work on the basis of the knowledge explicitly passed or made available to it, but also be able to dynamically acquire new knowledge through retrieval from external sources and possibly reasoning upon existing heterogeneous knowledge. In other words, the system should be able to react to the knowledge available to it by retrieving additional knowledge from which it can make further inferences. This is a necessary consequence of the limitations inherit in knowledge-based processing, which is that the process is only able to act upon the extent of knowledge available.

In a dynamically changing environment such as the Web, the domain of the content being processed can expand or change. In such a case, an IMMPS without dynamic knowledge integration will require constant human monitoring to ensure that its internal, closed knowledge base is regularly changed or updated.

The SRM proposes an architecture of expert modules (the components of the knowledge server) that supports the acquisition of additional knowledge from external sources through an "acquisition interface". The overall architecture of the expert module is, for the purposes of the model, introduced very generically. However it is clear that the acquisition interface shall provide a means for the retrieval of knowledge from other servers and its processing in order to be able to be added to the existing knowledge of the expert module and hence enable the inference of new facts as a result.

The Cuypers system in particular has been used as a basis for research to investigate the use of additional knowledge sources upon the multimedia generation process. An exploration of Semantic Web techniques [Geurts, 2003] focuses on the required knowledge that a system utilises to produce a multimedia presentation from an initial set of retrieved content. The approach explicitly models this knowledge apart from the system and makes it available to the generation process at runtime. It proposes a two-step, ontology-driven transformation process.

The first stage in the process organizes a set of media objects according to its knowledge of the domain referenced by their annotations. This domain knowledge is represented by a *Domain* ontology and the selection and organisation of the objects is controlled by the ontological relations of those subjects in the media objects' metadata. The implementation uses RDF annotations which contain instances of the ontological classes. Instances can then be associated through their respective class associations. However it seems this approach relies on the domain being known prior to system execution (i.e. the correct ontology is already pointed to from the system). As a result, a semantic graph is produced.

In the subsequent stage of the generation process there is an ontological model of the discourse knowledge, which allows the selection of an appropriate abstract presentation structure. The *Discourse* ontology is based on genres, each of which is associated with specific rules for the ordering, grouping and prioritizing of information. The association of media to a particular genre from the semantic graph is, by necessity, dependant on the prior knowledge of the domain. In other words, the discourse ontology is not entirely independent of the domain ontology. The resulting data structure is called a *structured progression* [Rutledge,2003]. It is this data structure which is used as the basis for the final stage of generating the multimedia presentation, guided by media design knowledge from a *Design* ontology.

Rather than hiding the knowledge required by an automated system to generate a coherent multimedia presentation implicitly in document transformation code, in this approach it is modelled explicitly using semantic techniques and so used by a process to produce a final presentation. The internal logic of the process can then be generic to the domain, discourse and design related aspects of the system decision-making as the actual decisions will be determined by the interpretation of the external knowledge. As the knowledge models in the research are RDF-based, this interpretation can be realised by RDF-specific query and inference languages.

The approach requires that the ontologies are available to the system prior to execution, rather than allowing for a dynamic acquisition of the knowledge once the domain is identified. The ontologies are also interdependent (the discourse ontology draws on the classes of the domain ontology, the design ontology draws on the classes of the discourse ontology) though there is support for the 'bubbling-up' of rules to permit re-use of more general rules where applicable. Regardless, this interdependency also rules out easily incorporating dynamic knowledge association into the system.

While this marks the use of knowledge acquired during the execution of the process, it remains a relatively closed system, no different from a system interpreting knowledge available at the start of execution. A particular domain is modelled and the discourse ontology is also related back to this domain. The interpretation of this knowledge is fixed by the implementation of the system. However, this is the only case in the IMMPS literature where knowledge is introduced to the multimedia generation process during execution, as opposed to being statically available prior to the execution. Even then, this work can not be said to demonstrate a truly dynamic incorporation of knowledge.

3.4.1 Summary

At present, the dynamic integration of further knowledge in the multimedia generation process is not a part of most IMMPS. This can be understood in the context that knowledge itself, even as the input to the system, is only considered in an implicit manner, i.e. the system expects a narrowly defined input from which it can derive knowledge specific to its understanding of that input. Dynamic understanding of some data relevant to the process, which would require the use of more general machine-interpretable formats is only really considered in the Cuypers system yet also in a rather narrow manner.

The architectural basis to support such dynamic incorporation of knowledge can be seen in the work on Semantic Web Services [Peer, 2002] which is in an early, prototypical stage. It is clear, however, that such services will be also of benefit to next generation multimedia generation systems.

SRM-IMMPS	All expert modules through an “acquisition interface”
Madeus	X
Cuypers	Some research in integrating RDF/OWL
MAVIS 2	X
MediaNet	X
Artiste	X
Shih’s IMMPS	X
MIPS	X
Otto et al	X

3.5 Presentation of multimedia on the basis of expressed concepts and their relationships

The final coherent production of a multimedia presentation depends not only on handling the syntactic issues inherent in the retrieval, adaptation and final format of heterogeneous content, but also the semantic issues, i.e. the higher level concepts which the media represents to the user. While low-level features are a common basis in an IMMPS for design decision-making, and are also inherently available to a computer system which can analyse such features, high-level features must be provided to the system, e.g. through textual annotation or Semantic Web metadata, as well as supported explicitly within the multimedia generation process.

The SRM notes that the Context Expert shall contain knowledge relating to “media objects and their semantic concepts”. It does not however produce any further guidelines for the representation of this knowledge, including the modelling of the media objects or their semantic concepts. However the availability of such knowledge is necessary if desired requirements of an IMMPS are to be supported such as personalisation to the user. The SRM itself notes as examples of this, “selecting the content of a presentation according to the user’s information needs, or when deciding which media to use so that a user’s perceptual abilities and preferences are matched”. This is only possible where the expressed preferences of a user (which are expressed as concepts) can be matched to the available media (which represent in varying ways those concepts).

In the MAVIS 2 system, a “Multimedia Thesaurus” was implemented. The initial MAVIS system excluded the selection of the spatio-temporal segments of media, which related to individual concepts, and was unable to link media objects together which represent the same concept in very different ways (e.g. from a different angle). To overcome this, a *semantic layer* was proposed which defines associations between the media representations available to the system and the real-world concepts that those media representations communicate to the user. This layer could support any arbitrary relations but was initially proposed to support common thesauri types – specialisation / generalisation and “related”. The MAVIS 2 system used these common thesaurus relationships to guide the presentation of the media retrieved in an user query: while the query results from the low-level analysis of the media would be ordered as in other multimedia systems, other media – determined by its concept being related to the concept queried upon – would also be presented in spatio-temporal relationship to the original result.

This approach is reminiscent of other approaches to knowledge modelling of multimedia, particularly in its conceptual division between subjects in the real world and the digital representation of those subjects. In the case of MAVIS 2, a knowledge model is explicitly used to construct a “Multimedia Thesaurus” and hence semantic techniques such as inference could be applied. However it is acknowledged that much depends on the initial population of the semantic layer [Tansley,2000b] (the issue of ontology building) and the correct association of the terms in the layer to the media. As low-level feature analysis has limited success in this field, the presence of suitable metadata with the media objects can aid immeasurably the process of populating the semantic layer. While it seems that the semantic layer has been realized in a proprietary application-internal way, its simple thesaurus structure could be realized by a standardized ontology language. The authors mention RDF, and an implementation using the RDF model would benefit from interoperability with other ontologies and the re-use of existing metadata. The major issue of the approach is the extendibility of the system to support further relationships. The demonstration system supports a simple set of thesaurus relationships and the system logic for the ordering of the media objects in the presentation component is clearly system internal. An external, declarative means would be necessary to easily alter the system to support other relationships.

A similar research project is MediaNet, where the conceptual relationships are expressed within a semantic network. The MediaNet system is a knowledge representation framework that uses multimedia content to represent semantic and perceptual information. It consists of conceptual entities, representing real world objects, and relationships between concepts. The concepts or relationships can be defined or exemplified by multimedia content. It also extends the set of relationships from traditional KR relationships such as generalization or aggregation with perceptual relationships based on feature similarity.

The MediaNet framework is comparable in functionality to an encyclopaedia or dictionary with a thesaurus. It is therefore similar to the Multimedia Thesaurus approach to MAVIS 2. While the latter restricts concepts to high-level, semantically meaningful real world objects (i.e. those referenced unambiguously by nouns in natural language), MediaNet incorporates perceptual concepts and relationships in the model, such as patterns and similarity matching.

While the MediaNet approach is seen as applicable to a broad range of applications that deal with multimedia at both the semantic and perceptual level, it is evaluated in the context of multimedia content retrieval. Because it maintains an association between the media in the process and the concepts in that media, as well as expressions of relationships between those concepts, it supports the final presentation of media according to concepts and relationships even though that is not the primary purpose of the system.

Both systems, MAVIS 2 and MediaNet, are using the conceptual expressions and their relationships within the context of retrieval, and the subsequent presentation of the retrieved content. The following systems are more focused on the actual realization of the final presentation according to concepts and relationships.

Shih defines a domain-independent architecture for a multimedia generation system which he calls an Intelligent MultiMedia Presentation System[28]. A multimedia presentation is modelled as multiple presentation windows organised in a tree structure, representing the linear realization of the multimedia presentation. Presentation windows are associated with facts and knowledge rules, which are inherited along the branches of the tree. The system also permits messages to be passed between presentation windows, which package parameters necessary for the knowledge rules. Given that this knowledge may be able to describe concepts and their relationships, the system is enabled to produce final presentations based on that knowledge. It does this by applying standard logical inferencing. The subjects and objects of the expressed facts are not a concern of the knowledge-based process, and carry no other meaning outside of their input to the inferencing logic. The variation in system-aware facts and rules is guided by the user interaction and the input parameters, in so far as they lead to alternative paths through the internal tree structure of the multimedia presentation process.

It allows component re-use and takes a procedural approach to presentation authoring. The authors give a few examples of implementations, demonstrating the systems application across different domains. Authoring requires the provision of: presentation resources, presentation knowledge, navigation rules and presentation

[28] Despite the same name, this work is not related to Bordegoni et al's Standard Reference Model for IMMPS

window layouts. A pre-requisite for this system is an object oriented multimedia database, meaning that all content must be integrated into a single source. It is however not extendible to integrate external resources automatically, or flexible to handle the same generation process in different situations i.e. there is no support for contextual information.

The MIPS system contains as its presentation component a “web builder” module. This module orders the retrieved media items in the final presentation according to the original concept, which determined their selection. The MIPS system retrieves information from multiple heterogeneous content sources and generates an integrated hypermedia presentation of the answer. The presentation system is tied to the HyTime model and does not include the temporal aspect. The knowledge base that is modelling the information domain uses a simple semantic network containing object sub-typing and instantiation, a “part-of” relation, a geographic “location-of” property and attribute possession and value. Again, the modular logic for processing these concepts and relationships into an ordered multimedia presentation is internal and hence not supportive of dynamic, flexible changes in the system.

The system from Otto et al is based around the use of an information model, which is built upon a core semantic structure by application-specific domain knowledge. The “information” is considered in terms of qualitative aspects (how useful) and characteristic aspects (how it has different uses or usefulness in different situations). Information objects are split into three top level sub-classes – thing, process and term. These classes can be further refined according to application need, yet in an application-dependent form. This knowledge allows for the definition of new classes and then the instantiation of those classes to produce the concepts. The results of the original information query can then be ordered according to the knowledge in the system of the concepts, which relate to the query.

The information space is effectively an ontology, in which information objects are instantiated and associated to media. Object-object relations are supported, as well as attribute assignment onto objects and their relations. The research literature also mentions ‘dependencies between attributes’ and ‘properties of ranges’ which suggests constraint languages and data type schematics. The conceptual modelling is an internal, proprietary approach which makes it difficult to integrate or extend with other knowledge models.

Common factors in these multimedia presentation systems reviewed here are that the conceptual model of the system is expressed internally and in a proprietary format, making this approach exclusive to the previous aim of dynamically incorporating external knowledge. The generation process, which is guided in its presentation decisions by this conceptual model, can not react to run-time changes which would rely on making updates in that conceptual model. The flexibility in which a presentation can be organized according to the concepts and relationships is also tightly constrained by the restricted, proprietary rules of the systems.

It is in the research work based on the Cuypers system that the most flexible approach to determining multimedia presentation by concepts and relationships in media can be encountered.

In this case, the research deals with inferring the spatio-temporal relations of media in a presentation from their semantic relationships as expressed in a RDF formatted annotation [Little,2002]. A service was built which queries a media archive whose content is marked up in RDF/Dublin Core (using the Open Archives Initiative guidelines [OAI,2001]). Semantic matching rules have been defined which map to spatio-temporal relations between media. The Cuypers system handles the overflow strategies i.e. the constraints that ensure the viability of the final presentation while seeking to respect the semantic relations as much as possible. The high-level architecture is shown below (Figure 3.4)

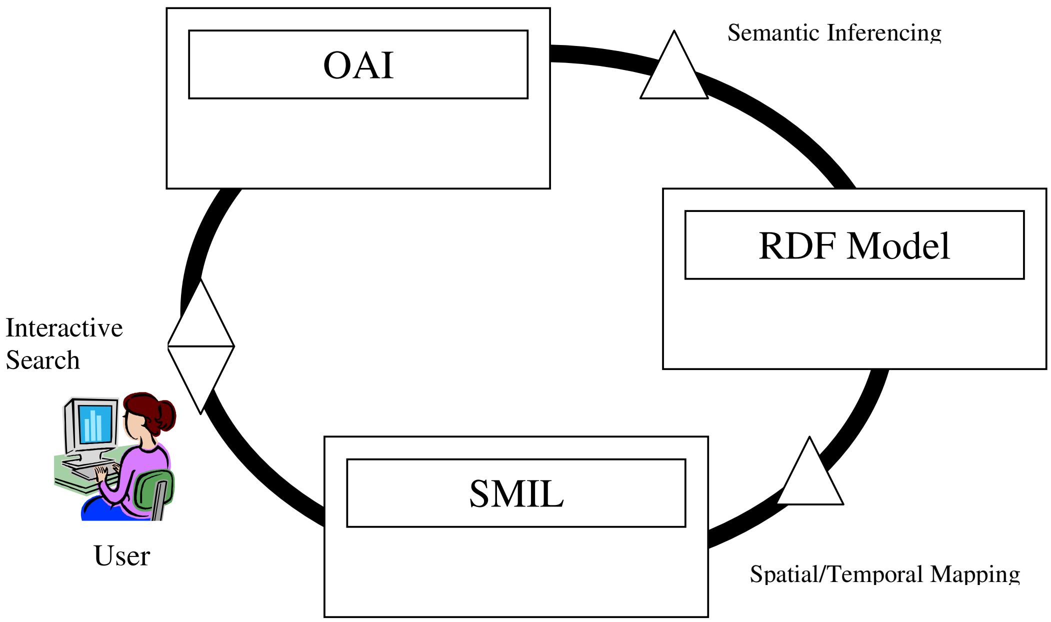

Figure 3.4 Architecture overview of CWI's semantic inferencing system

The multimedia generation process is realized in five stages:

- An *Iterative Search Process* across heterogeneous OAI archives, allowing users to select resources which directs the subsequent search focus,
- A *Semantic Inferencing* stage where the Dublin Core metadata is used to infer semantic relationships,
- A *Mapping* stage where those relationships are represented in the systems internal abstract presentation form as spatio-temporal relationships between media objects or abstract media objects,
- *Presentation Generation*,
- A user-directed *Presentation Regenerator* where the user may select a media object in the final presentation and hence repeat the process to generate a new presentation with that media object as the new focus.

The semantic inferencing is based upon the respective values of Dublin Core elements in groups of metadata entries retrieved from the database. For example, a *created* relation can be considered to exist between two resources if one resource's metadata identifies X as its creator, and the other resource's metadata identifies X as its subject. A top-level set of semantic relations were defined to enable consistent mappings, and these were taken from the MPEG-7 Semantics DS. Low-level, domain-specific relations can be defined for individual applications and their spatio-temporal mappings fixed by reference to their super-property from the top-level set.

Identified problems from this work were that the existing metadata (from the OAI archives) was inconsistent, sometimes inaccurate and based on simple resource discovery needs. A consistently applied normalisation of terms and an agreed level of detail in metadata is required if retrieval is to be possible and effective. Metadata also often lacked an explicit reference to the content it annotates, which is essential for automated retrieval. It was also found that the semantic inferencing rules were not universally true. There is a need for contextual validation. Contextualisation is not modelled in RDF/Dublin Core. Property values also lack unambiguous semantic meaning in the metadata model due to the representation of the objects of statements as literals rather than URIs drawn from a known schema or ontology, and the lack of data typing (e.g. so that dates could be compared). If semantic relations are to be inferable from heterogeneous metadata sources, there is the requirement that there is sufficient information available in the metadata and that this information is truly machine-processable in an automated system by reference to ontologies/semantic models which place the information within its domain.

3.5.1 Summary

Most of the systems reviewed here incorporate some means of ordering multimedia into a presentation through some form of consideration of what media signifies. Important aspects for this activity are an annotation of media with the concept it represents, properties expressing relationships between those concepts or concept types and rules for determining how these properties are represented in the presentation structure. However, concepts are typically internally identified, i.e. the identification is fixed prior to execution of the system and the system rules are written specifically in each case to use those identifications. Furthermore the properties expressed between concepts are limited and simple, e.g. thesaurus relations. Finally, the rules depend typically on the internal set of properties supported in the system and on relating to a specific multimedia model being used within the system and hence can not be interchanged between systems.

SRM-IMMPS	Foreseen in the Context Expert
Madeus	X
Cuypers	Mappings defined between RDF properties and presentation constraints
MAVIS 2	Fixed presentation based on thesaurus relations
MediaNet	Internally generated relations, fixed presentation
Artiste	X
Shih’s IMMPS	Internal knowledge model and rules
MIPS	Fixed presentation by small set of relations
Otto et al	Internal knowledge model and rules

3.6 Taxonomy

Eight systems in total have been reviewed in this chapter, representing research work in the field of knowledge based multimedia presentation systems, and compared to the Standard Reference Model for such systems, to which they are in varying degrees similar.

The table on the following page illustrates this review, indicating the extent to which the systems meet the requirements that were introduced from the first chapter. Empty boxes signify that the given system does not meet the given requirement. For the SRM we give the expert modules which are proposed to solve the given requirement, as for the systems themselves, we give a brief description of the extent to which they meet the given requirement, bringing together the individual summaries given in the earlier sections of this chapter.

3.7 Conclusion

This brief review of multimedia presentation systems, which incorporate as part of their research a knowledge-based component, with respect to the requirements presented from the motivation of this thesis indicates that current system implementations do not satisfactorily meet all those requirements. In contrast, the requirements are all within the bounds of the SRM for intelligent multimedia presentation systems.

Intelligent multimedia retrieval is enabled and enhanced through a conceptual model of the multimedia content domain, yet retrieval from conceptual models has continued to base itself on low-level features, whether perceptual features of visual or audio media, or textual string matching.

The processing of different data formats and structures is also primarily fixed within the systems, while operating on heterogeneous sources will necessarily require flexibility in data processing. This calls for dynamic interpretation of the retrieval data formats and structures by the system, and the ability to handle both syntactic and semantic data.

The contextual adaptation of multimedia presentation may also benefit from a knowledge modelling of that context. However, while content domain models could conceivably be designed for specific system applications, context is a universally shared domain and requires shared, standardized conceptualisations. The reviewed systems do not show clearly how they might interoperate with such standards as might be encountered on distributed networks such as the Web. Rather the conceptual processing remains a low-level exercise (e.g. string matching), when a high-level approach (e.g. the use of URIs as in the Semantic Web) would be more effective.

Intelligent Multimedia Presentation System	Retrieval of data from different sources	Processing of hetero-geneous data formats	Incorpor-ation of contextual adaptation	Dynamic integration of external knowledge	Presentation based on concepts and relationships
SRM-IMMPS	*Application Expert*	*Application Expert*	*User, Device, Context Expert*	*All expert modules through an "acquisition interface"*	*Context Expert*
Madeus		use of XSLT so limited to structured XML data	XSLT for additional adaptation		
Cuypers			Context as additional input to process, handled internally	Some research in integrating RDF/OWL at start of process	Defined mappings between RDF properties & presentation
MAVIS 2					Fixed presentation based on thesaurus relations
MediaNet					Internally generated relations, fixed presentation
Artiste	RDF query mapping to SQL	Limited to relational database schema			
Shih's IMMPS					Internal knowledge model and rules
MIPS	Internal query mapping				Fixed presentation by small set of relations
Otto et al			Internal proprietary context and rules		Internal knowledge model and rules

Another factor with contextual processing is that it draws upon knowledge beyond that of the multimedia content domain. It raises the requirement of dynamically integrating additional knowledge during the multimedia presentation process. However the reviewed IMMPS do not satisfactorily take this into account.

The final multimedia presentation will be coherent in structure when layout rules are successfully applied to meet constraints, yet only coherent in their semantics when layout can also reflect the concepts represented by the media and their relationships to one another. The introduction of knowledge-based approaches to multimedia presentation systems should support this semantic coherency yet systems remain primarily syntactic in their implementation of the multimedia generation process. This may be a result of semantic approaches being seen as immature, or the lack of overlap between the AI and the multimedia communities.

Having seen how the presented requirements are not met by existing IMMPS work yet are within the bounds of the Standard Reference Model, we turn to our work in this field, proposing and implementing a new approach to IMMPS which can meet all these requirements.