# Denotative and Connotative Semantics in Hypermedia: Proposal for a Semiotic-Aware Architecture

*Frank Nack and Lynda Hardman*

CWI Amsterdam
Kruislaan 413, P.O. Box 94079
1090 GB AMSTERDAM, The Netherlands
E-mail: {Frank.Nack, Lynda.Hardman}@cwi.nl

## ABSTRACT

In this article we claim that the linguistic-centred view within hypermedia systems needs refinement through a semiotic-based approach before real interoperation between media can be achieved. We discuss the problems of visual signification for images and video in dynamic systems, in which users can access visual material in a non-linear fashion. We describe how semiotics can help overcome such problems, by allowing descriptions of the material on both denotative and connotative levels. Finally we propose an architecture for a dynamic semiotic-aware hypermedia system.

## Keywords

Hypermedia, hypertext, multimedia, denotation, connotation, semiotics, ontology, conceptual information space.

## 1 INTRODUCTION

For domains with a critical amount of media cross referencing, such as the disciplinary context of theory, history, and anthropology of film, hypermedia offers ways to go beyond the traditional mode of exchanging information. Prior to the use of hypermedia, researchers in that domain presented their work using a linear textual representation in which they refer to the audio-visual material they analyse on an abstract verbal level or use embedded audio-visual surrogates, such as key frames for video, or scores for audio tracks.

The core problem of this approach is that the totality of the information is carried by three structures, namely the text, the image and the film the image represents[1], of which in the best case two are represented, specifically the text and the image. Both the text and image are contiguous in their message spaces and hence they can be co-operative. However, since the carrier of the message is either made of words (text) or lines, surfaces, colours, and shades (image), both structures remain separate [5, p. 16].

Of the two structures, that of language is well suited to suggest associations or ideas about the visual material and perform a strict analysis on a semantic level. However, by doing so the text functions as a denominative anchor that selectively reveals not the totality of the visual continuum but only a fraction of its signs by recourse to a taxonomy. Thus, the traditional linear and text-centred way of discourse about still and moving images can never fully capture the full diversity of visual material.

Now, using hypermedia, researchers are able to replace the manuscript with a conceptual information space that permits a parallel comparison of a theoretical train of thought with the full actual media data. Moreover, the conceptual information space also allows the simultaneous critical revision of the specified aspect provided by other researchers.

Such an environment requires that the information units buried within the established relations between the single media pieces be made available. Moreover – and this is even harder to tackle – discrete information units hidden in the structure of the individual text, image, video, audio or tactile entity need to be made available, so that new compositions of various elements can result in the establishment of new meanings. When this is achieved, we are in the position to study each medium exhaustively and understand the manner in which they complement one another.

The result is an information space in which a community as a whole can develop and strengthen its own knowledge and practice or, in other words, provide through the information space its perspectives on the domain. It is this process of 'perspective making' [9] that underlies the building of a community's identity, e.g., its basic assumptions, goals, terminology, and modes of discourse.

The ideal situation would be to make such an information space not only accessible for the domain community but also to other interested parties, such as the general public, either for pleasure or for educational purposes in the context of life-long learning.

The problem we are then faced with is that of 'perspective *taking*' [9], which refers to the process of trying to engage with

[1] In fact the image performs here the same task as a textual description of an image: it describes in an imprecise and incomplete way, it changes structures and signifies an impoverished subset of what is shown.

another community's perspective. This can be extremely difficult when their respective ways of knowing assumes different agendas or does not match at all.
A potential way of bridging the gap between various viewpoints on the analytical and perceptual side of information is the creation of an adaptive hypermedia environment. This would be able to facilitate the development of different perspectives based on computer-supported analysis of primary multimedia sources combined with expert opinions and an adaptive user model.
Whether for technical or economic reasons, the current practice in hypermedia applications and systems is still to provide audio-visual information primarily for descriptive purposes while the main information is conveyed through text.
While this methodology provides a means for modelling argumentative discourse in general [16], sophisticated requirements for scholarly argumentation [31, 28], or establishing large narratives [62], we claim that within hypermedia systems this linguistic-centred view needs extension through a semiotic-based approach if different media are to achieve real analytic parity with text.
Semiotics is here understood as the study of signs, in which linguistic and cinematic signs constitute a specific topic [21, pp. 591 - 593]. We argue in this article that a semiotic-based methodology facilitates support for adaptive [15, 20] and adaptable hypermedia [52] in its attempts to dynamically manage the complex task of hypermedia presentation and navigation.
In this article we first discuss the aims of research in the theory, history, and anthropology of film. We then present the principles of visual-based[2] semiotics and demonstrate the use of these principles on a hypermedia demonstrator for film analysis. Finally, we describe a potential semiotic-based hypermedia architecture and conclude with future directions for this research.

## 2 CONTEXT

Like most research in arts and humanities, investigations on film follow an interpretive, associative method based on historic-cultural materials, including primary sources as well as secondary materials. The phenomena to be analysed can be categorised as follows [2]:

- The raw material: includes questions about the medium and its relations to reality, photograph and illusion, its use of time, space, colour, sound, props, actor make-up, etc.
- The methods and techniques: includes questions about the creative and technological processes which shape and treat the raw material, as well as the underlying psychology or economics.
- The forms and shapes of film: includes questions about film categories, the adaptation of other art forms, genre and audience expectations.
- The purpose and value of cinema: explores the goals of cinema for humankind.

Thus, researchers of this domain work within a theoretical system that provides information on the relationship between audio-visual units (comprehension) and the ideas they represent (interpretation). The comprehension/interpretation couplet stands for the evolving process of making sense of a film, where various categories of meaning compete, such as [10, pp. 1 – 18]:

- Referential meaning: The spectator/critic constructs the film world by drawing on filmic and extrafilmic conventions, conceptions of causality, space, and time and on knowledge about concrete items of information.
- Explicit meaning: The viewer assigns a conceptual meaning to the story.
- Implicit meaning: The viewer assigns themes, problems, issues or, in other words, a symbolic quality to the film.

The relevant point is, that the various categories are used in a functional and heuristic way, as assumptions that generate hypotheses about particular meanings and thus result in a cycle of meaning production. The main objective of research in film theory is therefore to use assumptions, expectations, and exploration for arranging highlighted cues as the basis of a critical inference that explains how the visual expression is reconstructed into something meaningful by the perceiver's presuppositions at the time of perception [3] along with the various legitimated codes and sub-codes the receiver uses as interpretational channels [21, pp. 593 - 598]. In other words, the system of film theory facilitates the comparison and contrast of distinctive views and approaches with our own.
However, it must be made clear that the main media in research on film is visual. If we wish to treat visual media on an equal basis as text in a hypermedia-based discourse, we have to investigate its means to elucidate, sublimate, or rationalise information. When this is achieved we will have a better understanding of how text and visuals complement each other. Thus, we now examine briefly the semantic structures in an image and their variable nature depending on the sequence images appear in.

---

[2] The discussion in this article will focus on visual material. The authors are aware of the fact that the absence of sound is a strong qualitative drawback, since sound provides three out of the five standard categories for substance and is, therefore, a powerful commentative story element. Nevertheless, sound was excluded from this discussion because its integration is too great a challenge for a paper of this size.

# 3 VISUAL SIGNIFICATION

Unlike spoken or written language, visual material, in particular the photographic image, *depicts*, and the viewer does not usually have to struggle to identify what it shows. The denotative power of film, the optical pattern, communicates a precise knowledge, which releases the audience from the process of decision-making but nevertheless leaves a problem of *interpretation*. The process of interpretation is in particular important because we can see an image in two ways – in isolation or in context. An image shown in isolation is a form of utterance that provides an identifiable semantics. The same image presented in a sequence might appear with a modulated semantics because the order created new levels of meaning. The same effect appears in sequences of image sequences, which we call film. Since the relationship between the two representational systems, i.e. the image (space) and order (time), is complex it is useful to quickly examine them separately to determine their relevance to the generation of meaning. In the following sub-sections we thus discuss the image, the shot, montage, and the rhetoric of a sequence.

## 3.1 Image

If we look at an image we first perceive it on an optical level where we try to identify as many objects as we can in the available time of perception. Each object is mentally transformed into an iconic sign. An iconic *sign* is usually described in semiotics as a sign processing some properties of the object represented [49, p. 2228]. Eco showed extended this generalised view by showing that the signification of iconic signs is based on a socially determined reticular system of small semantic systems (codes) and rules for their combination. He defines a number of codes, including [21, pp.596 – 598]:

| | |
|---|---|
| *Perceptive codes* | establish the conditions for effective perception. |
| *Recognition codes* | Structural blocks of perceptive qualifications (signifieds), according to which we recognise objects or recall perceived objects, such as black stripes on white fur, based on which objects are recognised. |
| *Tonal codes* | Systems of optional variants (prosodic features) already conventionalised. They accompany the elements of iconic signs as an added or complementary message. |
| *Iconic codes* | Perceivable elements that can be subdivided into *figure*, *sign* and *semes*.<br>A *figure* forms conditions for perception, such as relationships between object and background, contrast in light, geometrical proportions.[3]<br>A *sign* denotes, using conventionalised graphical methods, units of understanding (nose, ear, sky, cloud), abstract models, or idealised diagrams of the object (the sun as a circle with thread-shaped beams).<br>*Semes* are complex iconic phrases, such as ‘this is a man standing in profile’. They are the most simply catalogued and since the iconic code works most often on their level only, semes are the key factors in visual communication.<br>Iconic codes change readily within the same culture, due to their contextual interlacing (a horse as part of a shop label may suggest the availability of equestrian products, while a horse on a traffic sign may suggest "beware, horses on road"). |
| *Iconographic codes* | connote more complex and culturalized semes that are, despite their complexity, immediately identifiable and classifiable, such as "the four horsemen of the Apocalypse". |
| *Rhetorical codes* | Models or norms of communication, which can be divided into *rhetorical figures* (e.g. *metaphor*), *premises* (e.g. a man riding along a never ending prairie can connote loneliness), and *arguments* (which create connotations based on the succession or opposition of different images). |
| *Stylistic codes* | A stylistic feature, such as the mark of an author (for example, a man walking along a road tapering off into the distance suggests “Chaplin”), or the typical realisation of an emotion (a woman who leans seductively against the curtain of an alcove suggests “Erotic of the Belle Époque”) or the typical realisation of an aesthetic, technical-stylistic ideal (as in cubism, where objects are portrayed in abstract, geometrical forms). |

Later in the paper we will see why these codes become important in the generation of hypermedia presentations. For the moment it is important to show that the iconic code is by far the most valuable code, because it defines the articulation potential of visual material. The creation of meaning in visual material is based on a *triple articulation* of figure, sign and semes and receives its expression by convention. The fact of a conventionalised triple articulation is important since it describes the essential difference from natural language, which has two articulations (phonemes and morphemes). Thus,

---

3 All of these codes have been developed and refined by other visual arts, i.e. painting, sculpture and photography. Arnheim [3] proposes ten determinants: balance, shape, form, growth, space, light, colour, movement, tension and expression.

comparing an object in an image with the corresponding word, the visual object always exceeds the concept of the word, as the image will portray specific qualities about the object for which the word is simply inadequate.
The organisation of signs in an image is provided by syntagmatic and paradigmatic structures [30, p. 74]. The syntagmatic structure represents a sequence of signs in which the relation of parts determines their meaning. Figures in the iconic code, iconographic codes and arguments in rhetorical codes represent syntagmatic structures. A paradigmatic structure represents potential substitutions in which a range of potential candidates can take the place of a sign in the syntagmatic structure[4]. Signs and semes represent paradigmatic structures.
It should be stressed here that the diversity of the semantic systems (codes, syntagma, paradigm) provide, with their combinatorial possibilities, the foundation for a subjective interpretation by each viewer, as mentioned earlier. Codes, however, can only realise their full potential impact if there is an awareness of them and if they can relate to existing knowledge.

## 3.2 The shot

Of higher complexity is the relationship between an image, or frame as it is called in film, and the shot, where a shot is "a single piece of film, however long or short, without cuts, exposed continuously" [39, p. 452]. The significant additional element here is time, which provides the basis for the understanding of action, distance and the relationship among characters, based on the relationship between frames within a shot and their rhythmical variations.
The compositional use of *focus*, for example, through which the foreground, middle ground or background are emphasised, guides the perception of a shot. If all planes are represented in focus, they are attributed with the same level of importance, whereas emphasis can be achieved by use of focus for a part of a frame. *Citizen Kane*, by Orson Wells, provides many well-known examples of the use of focus in these ways. Of even stronger impact than focus is *camera movement* around the imaginary vertical axis (pan), the horizontal axis (tilt), and the longitudinal axis (distance from lens to the subject). The *tilt*, for example, presents the eye-level from which a scene is perceived and thus can affect the importance ascribed to an object (for example, high-angle shots may diminish the perceived importance of an object).
Using the dynamic qualities of film, specific elements can, in one shot, directly provoke an emotional reaction. Imagine a shot in which the camera follows a character through a group of cheerful, passionate people. The appearance and disappearance of the group in itself can suggest the character's sense of isolation.
The *tempo* of a shot can also provide information. The intense feeling of fast movement may excite, while calm movement expressed, for example, through the slow rolling of waves filmed from a static camera position, may encourage feelings of relaxation. Related to tempo, is the *perceived duration* of the shot. The actual duration of a long shot full of people and action may well be identical to one of the close-up of a face, and yet the latter will be perceived as being longer. Hence, the organisation of perceived duration is more complex than the actual duration of a shot[5].
It should be made clear that, in itself, the shot is an individualised unit with an invariant semantics based on the combination of the components in the frames (see [21, pp. 601 – 604]). As Eco pointed out, we are confronted again with a triple articulation, only that here the iconic sign takes the role of the basic unit (kinesic figure). Within the temporal continuum such a sign is first of all meaningless – the image of a car in a frame whether it is moving or not and if in which direction. Kinesic figures only join together between frames in the flow of motion, to form kinesic signs (similar to the iconic sign, which is static). Analog to iconic signs, kinesic signs can build up complex statements (kinesic semes), which are comparable to a visual utterance made up of a number of movements. The combinatorial possibilities are limitless.

## 3.3 Montage

The final level of generating meaning with visual material to be considered is the way in which content of a shot can be affected by other shots, which is the domain of *montage*. Montage is based on two distinct, but mutually influential, aspects of our understanding of film:

- The meaning of a shot depends on the context in which it is situated;
- A change in the order of shots within a scene changes the meaning of the shot as well as the meaning of the scene [32, pp. 52 –53; 22, pp. 33 – 58; 23, pp. 11 – 57 and pp. 327 - 399].

---

4 A particularly interesting point made by Jakobson is that a sign system does not consist only of the two fundamental structures (paradigmatic and syntagmatic), but that each crystallises into a rhetorical device, i.e. the paradigm into the metaphor and the syntagm into the metonym (extending Saussure's syntagmatic and associative understanding of the linguistical planes of meaning, as described in Saussure [55, p. 123]). This means that even these "free" variations deal with codes that are based on systems of opposition and difference within the language of a culture, a social group or an individual.

5 For a discussion of the interesting relationship between rhythm and shot composition, see Eisenstein's article *Vertical Montage* in Eisenstein [23, pp. 327 – 399], which provides diagrams which describe a sequence of his film *Alexander Nevsky* in musical terms; see also the first two chapters of Burch [12, pp. 3 – 31], which feature the use of analogy between serial music and montage.

Gregory [25] is responsible for a detailed analysis of the importance of context and order in film editing. Gregory claims that not every combination of shots creates a meaning, but there are restricted conventions that can help create larger meaningful entities. His key elements for creating meaning by joining shots are *assertions* and *associative cues*.

An assertion is the relationship between two elements. There are many different types of such relationships. For example, the description of an attribute (such as *red* for a car) could be as important as a simple action (two men shaking hands). Gregory argues that a given shot "A" can build divergent assertions with other shots by using various subsets of the information gathered from shot "A". This is especially important, as it means that the juxtaposition of shots can be analysed, in that the shot can be used as a variable collection of information rather than a fixed visual description.

Associative cues result from the combinations of the indicators that make the creation of meaning possible. Gregory introduces two main groups of cues as being important in the creation of assertions. The first includes cues for the surrounding space. Most human activities, human roles or objects are associated with specific locations. The conceptualisation of space is, therefore, an elementary principle of the analysis and organisation of material in editing. The second type of cue is related to human actions. Thus, montage establishes meaning by building up sequential structures.

### 3.4 Sequences

The sequential structures built up through montage provide the complex and intricate syntax for film narrative. The temporal aspects of a film can reinforce meaning. For example, in the film *High Noon*, the real time of the film emphasizes the structure of the sequences and thus their tempo and shifts. Tarantino's *Pulp Fiction* is an example of the exact opposite of this rigid structure. Here four stories are combined together, and over time (here at crucial points of the narrative) the seemingly disorganised pattern falls into place. The repetition of shot devices can also serve to reinforce meaning. A film composed mainly of close-ups excludes information about its setting and becomes claustrophobic, whereas a predominance of long shots emphasises context over characters.

There is of course a temptation to synthesise various theories of montage into a logical pattern that should reveal the rhetoric structures within film. The best-known approach was the "Grande Syntagmatique" by Metz [36, pp. 108 – 146], in which he describes a system of binary oppositions indicating how eight types of montage (autonomous, chronological, descriptive, linear, continuous, and organised) were connected logically. However, the above description of meaningful structures in visual material leads to strong reservations about the approach of describing a film through its syntax, a reservation which is partly shared by Metz, who asserted that the syntax of a film is understood because the film has been understood, and only when it has been understood.

Looking again at Eco's triple articulation we see that its important point is not to identify a fixed number of articulations in a fixed number of relationships. After all an information unit (iconic or kinesic) can appear in any articulation, depending how the viewer looks at it. Thus, Eco's approach is much richer than the one provided by Metz, facilitating not only the representation of frame intrinsic information allowing the identification of the invariant meaning of a shot, but also the variable semantics of a shot, depending on the context it occurs in.

### 3.5 Consequences of visual signification for hypermedia

The work of Eco [21] showed us that the meaning of visual material is arbitrary and conventional, based on a myriad of perceptual, cognitive and cultural codes embedded in it. Thus humans have to learn how to interpret them on various levels of detail. This process of interpretation varies from individual to individual.

Moreover, we saw, through the work of Eisenstein [23], Kuleshov [32] and Gregory [25], that visual signification, though based on common human knowledge and thematic structures, provides its own temporal-spatial realities based on patterns of juxtaposition which are interwoven in the narrative structure.

For the representation and use of visual media in dynamic digital environments this means that, due to the triple articulation of visual material, we have to represent both its denotative and connotative aspects. It is clearly the connotation that relies on the denotation. This means that we can use the denotative representation to analyse, interpret or generate connotations on visual material but we cannot deduce denotative aspects of visual material on the mere basis of connotative descriptions.

Based on the above discussion we can now return to our original starting point: how do we facilitate the presentation of a discourse in a hypermedia environment where the essence of the argument is hidden in the relations between and the contents of various media.

## 4 SEMIOTIC-BASED HYPERMEDIA – A SCENARIO AND ARCHITECTURE

The limitless ways of perceiving visual material results in a critical discourse in the disciplinary context of theory, history, and anthropology of film, that establishes over time a knowledge space of collective interpretations This allows, by following the branches of interdependencies, a comparison of the most diverse theories, originally based on different perspectives. This knowledge space of various media is very much what Barthes describes as "… the metaphor of the text is that of a network, if the text tends itself, it is the result of the combinatory systematic…" which can thus be "…. experienced only in an activity of production" [5, pp. 157 – 161].

This notion of a semantic-network based discourse was from early on an objective within hypertext research [18, 27]. Since then we have seen further developments on modelling argumentative discourse in general [16], sophisticated requirements for

scholarly argumentation [31, 2813,14], or establishing large narratives [62]. An important role in these approaches plays the role of linking, as the mechanism to represent the dynamic and rhetoric of hypertext, a theme common in hypertext literature [6, 34, 40, 33, 60[6], 70[7]]. In fact, some work describes in some detail the relation between the structure of hypertext and film, which is based on a simple equivalence of nodes to shots, and links to edits [37, 35].
In particular Miles' article [37] is of interest, not only because he organized it as a hypertext allowing moves from hypertext to film (in the form of QuickTime clips) as simply as shifting from an abstract academic argument to a personal comment. It is of interest because it compares the major narrative structures of cinema, based on the work of Metz [36], with those of hypertext and shows clearly that both work is similar on a syntagmatic level, where the larger context of the narrative helps to determine what a particular link, or sequence of links, might mean. Mancini's chain of argument and theoretical deployment follows a similar path [35]. Both basically argue that the nodes in the hypertext structure need no explicit classification but instead can be left open so that their roles are defined depending on the relations connecting them. (see [13, 14] for the same approach for scholarly discourse). A similar train of thought can be found in work where the nodes exclusively contain video material [8, 17, 38, 56, 53].
We argue that the approach described and applied by Miles [37] works well because it uses a link taxonomy for its argumentation structure (canonical, commentary, quotation, reference, external) in a medium that in itself is structured on the same abstract connotation-based level, that is text. When the information node, here a page, turns into a composed entity of various media, however, the established structure starts to disintegrate.
For example, a node offers a commentary link titled "rescue" which turns out to be a composed node with textual and visual information about a sequence of Keaton's *Our Hospitality*. The total composition of this node, such as described in FIG 1, supposes to be the comment on the point raised in the previous node. A first quick view already tells us that of course not everything is part of the comment, such as the iconic bar at the top, which defines the colour code[8] for the link taxonomy, or the top title, which explains the conceptual ground we are manoeuvring in, that is the network provided by this hypertext. Thus, it takes some time to identify that the comment on the argument on the previous page is established here by the video clip and the text.

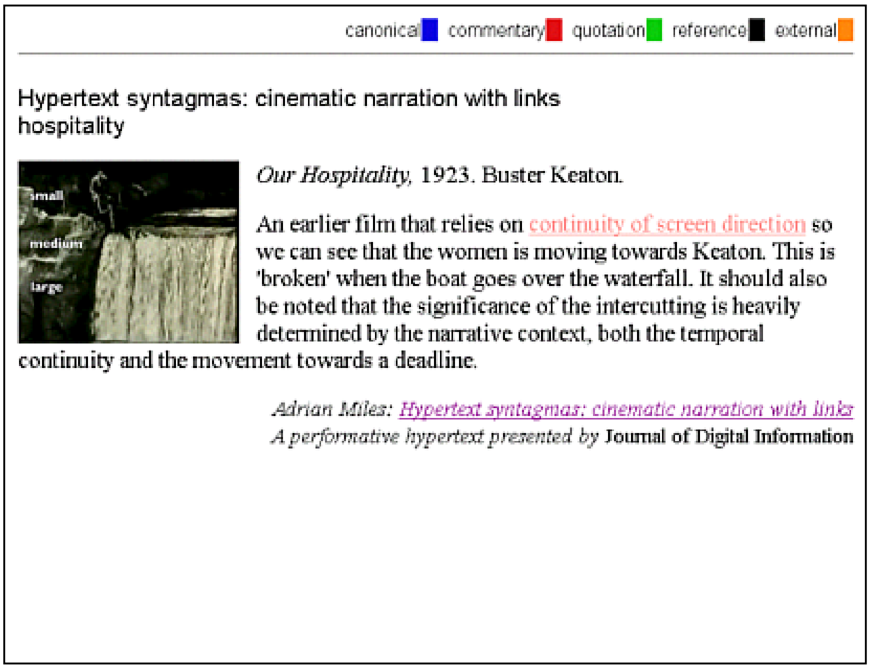


FIG 1: One node in the hypertext 'Hypertext Syntagmas: Cinematic Narration with Links" by Miles [37].

However, what is the relationship between the two items? There is no explanation available, such as the icon bar at the top. We can assume that the clip might be more important because it is further to the right than the text, but it could also be that the text is revealing something the video is unable to explain. So we might view the clip before reading the text, not really knowing what we have to expect. While watching, we will understand that the piece is part of a story, but we have no idea of

6 It should be mentioned that besides the described theory on semiotics in film, there are other approaches towards a methodology for discovering the stylistic strategies of a cinema of poetry (see [36, p. 81; 48]).
7 This reference is a representative for attempts in hypertext to animate link structures or nodes.
8 Links to the canonical text are blue, links to quotations are green, links to additional commentary are red and links to the references page are black. Traversed links appear lighter in tone than unvisited links.

its context, because we might not know the film. As a result, we might think about the relationship between the two characters and miss the point of the comment in the first screening. However, it might be possible that we first read the text, unaware of what the various points made in the explanation refer to, since we only see a static frame. At the point of watching we would know what we have to look for – and still miss the point because we were either completely taken by the drama happening in the clip or annoyed with it because we expected to see something different. Perhaps we read the text and simply do not see the relationship between explanation and material, because we are unable to identify the cutting.
The above example is taken from a relatively simple information network of approximately 200 nodes of which a great number are complex in their internal structure (text nodes containing various rhetoric structures) or compositional structure regarding the use of media. Miles himself assumes that the text cannot be read exhaustively and thus he provides on the cover page, a list of canonical links that branches to those nodes with the kernel of the argument.
Our aim is to support large hypermedia knowledge spaces. These spaces will mainly grow due to the work of specialists into environments with large numbers of nodes and relationships. Moreover, we wish to allow non-experts to access these rich information sources. For that we can build on work such as discourse modelling [13, 14] and hypertext rhetorics and presentation [37, 35, 70]. However, due to the complexity of the information space it is important to provide experts and non-experts with a means of accessing and evaluating the relevant material. Since neither the individual user requirements nor the requested material can always be predicted in advance, we claim that a system must be provided with knowledge of simple codes, i.e. collections of objective measurements for media units representing prototypical style elements, which are combined with high-level conceptual descriptions supporting contextual and presentational requirements. Using such a combinatorial approach, it is possible to establish conceptual presentations that support a better understanding of the topic, so that the system can find satisfactory solutions for upcoming questions (e.g. based on the content of a film sequence), misunderstandings (rearrangement of presented material) or non-understanding (creation of new argumentation sequences).
In the next section we introduce a demonstrator for a large information network on the director Sergej Eisenstein to clarify the required structures for content representation, link representation and argument presentation.

## 4.1 The Ecoian Eisenstein

The work of Eisenstein provides an excellent test case for being represented in a large knowledge space. His ideas concerning the construction of meaning in cinema, mainly based temporal and spatial problems within montage, addressed in his media theory as well as his practical film work, and his associative and thus often contradictory way of analysis make it nearly impossible to present his work in a traditional linear way.
The approach we envision to facilitate access to his work on various expertise levels is to transform his work and work referencing it into meaningful units and reconstruct them into a network which not only allows an understanding of the various ways Eisenstein addressed cinematic problems but also how these ideas influenced others.

### 4.1.1 Development of the idea

We deconstruct some of Eisenstein's theoretical articles in the literal sense of the word. We copied the text onto paper and cut it into pieces on the level of a paragraph and sometimes of a single sentence. Each piece was labelled with its origin (the original text it was taken from), classified with the role it played in the particular argument (e.g. a definition, a comment, a conclusion, etc.), and additionally annotated with information about the author of the piece and the date of production.
Our particular interest was focused on those information entities where a reference occurred and how these items contribute to the overall discourse. Physiscal references of the sort "… (when the Potemkin sails unharmed through the admiral's squadron)…" [23, p. 51] could be easily transferred into a representation of coordinates of the identification of film source, and frame specification of the relevant sequence).
More difficult are the various referential possibilities on a sub-level, in particular if the reference to an audio-visual source is provided by another audio-visual source. A typical example is demonstrated by the station scene from De Palma's 'Untouchables' and the arrest scene in Gilliam's 'Brazil', which both refer to the 'Odessa steps' scene in Eisenstein's 'Battleship Potemkin'.
Where de Palma refers to one episode of Eisenstein's complex montage piece (the part of the baby carriage), Gilliam uses a number of parts but refers to them either in form of nearly identical shot composition (the marching soldiers), or in a more subtle way, by referring to a well known scene (again the baby carriage) where the carriage is replaced by a hoover (which makes sense at this point because Brazil is a comedy). Moreover, where the visual reference in de Palma's film supports the story development only on a suspense level (will the baby be rescued during the fight between police and gangsters), the references in Gilliam's version not only support the story at this point (will Archibald Tuttle's group of terrorists be able to save Sam Lowry form the powerful state organisation 'Information Retrieval') but also fits on a meta level, since Gilliam's film is, very much like Eisenstein's 'Battleship Potemkin' working on themes, such as brotherhood, terrorism, liberty, torture, bureaucracy, freedom, and totalitarism. Thus, what is required here, is a reference system that not only allows the reference on object level (i.e. mark a region in a film that represents the relevant objects or film technique, and generate a link to the source these items are referring to) but we also need an environment that allows the connection of references (on the various code-levels as described earlier in section 3.1) which facilitates the recognition of complex references and hence the comparison of referential systems.

### 4.1.2 Network structures

Incorporating the results of our findings with the research on cinematic hypertext structures [37, 35] and work on discourse-based hypermedia systems [13, 14], we suggest the notion of semantic networks, a particular form of knowledge representation [11, 57]. FIG. 2 provides an overview of the structural elements in the network. At this point it is useful to examine the main features, i.e. nodes, relations and anchors, in further detail.

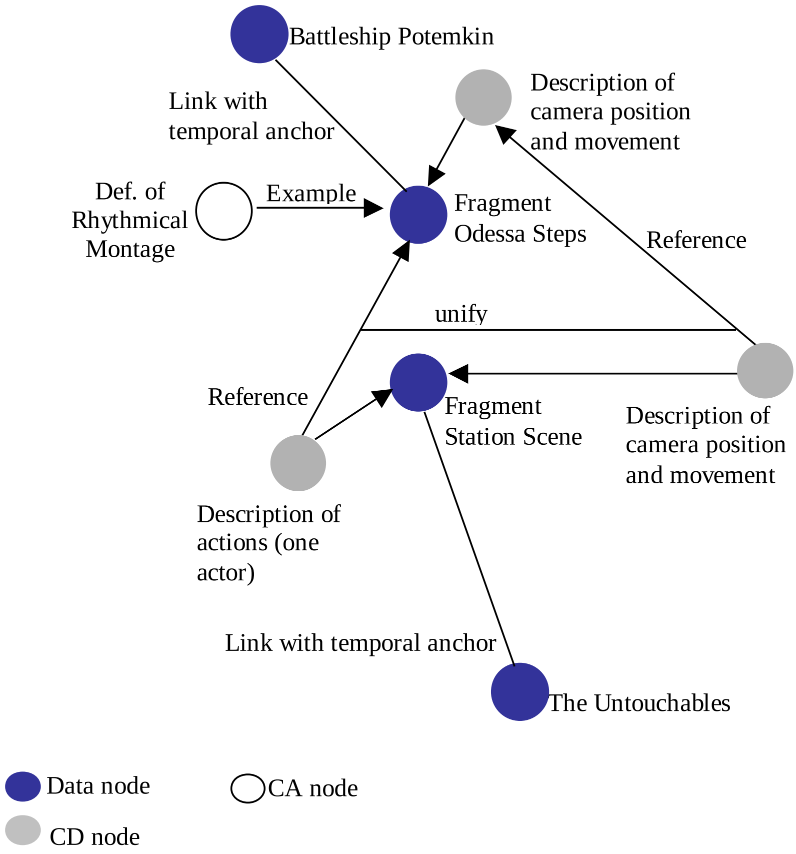


FIG 1: Graphical representation of the essential elements in the suggested hypermedia-based information network

We distinguish, in an abstract way, between three types of nodes: data nodes (D-nodes), content description nodes (CD-nodes) and conceptual annotations (CA-nodes).

A D-node represents physical material of any media type, such as text, audio, video, 3D animation, 2D image, 3D image, and graphic. The size, duration, and technical format of the material is not restricted, nor are any limitations present with respect to the content, i.e., number of actors, actions and objects. In FIG. 2, we can see that a data node might contain a complete film, as conveyed by the nodes 'Battleship Potemkin' and 'the Untouchables', or merely a scene, as represented by the nodes named 'Fragment Odessa Steps' and 'Fragment Station Scene'. A D-node for text might thus facilitate the identification of the document or a unit in a document – down to a word. For D-nodes representing audio-visual material we require a representation as first established by Bloch [7] and then fully explored by Parkes [47] to allow for '…minimal described unit of film sequence at the level of events i.e. the constituent below which descriptions, at the level of events, are not attached.' [47, p. 44]. The largest unit addressed could be a film, the smallest would be a single object within a frame.

CD-nodes and CA-nodes.are best understood as instantiated schema providing either denotative, characteristics of the data (CD-nodes), or connotative material (CA-nodes).

The important aspect for the relation of D-nodes to the other two type of nodes is that we allow the notion of multiple partially overlapping annotations of CD-nodes and CA-nodes on D-nodes, as suggested by Aguierre-Smith and Davenport in their Stratification System [1]. This facilitates the requested flexibility, extensibility and dynamics of scholarly work in the domain of theory, history, and anthropology of film. Interpretation notes are necessarily imperfect, incomplete and preliminary, because they accompany and document the progress of interpretation and understanding of a concept. Thus, any aspect of cognition might be illuminative for other research interests and should remain accessible. Annotating is dynamic and iterative work that maps the, not necessarily structured, process of the comprehension of a concept.

The use for denotative description of visual material, in particular with respect to its retrieval and reuse, is well represented in research on knowledge representation in AI [7, 47, 1, 19, 41]. In particular the work by Davis [19] and Nack [41] with respect to the representation of mise-en-scene (period and location), characters and objects (physical description as well as position in the frame), actions, and cinematography (lens and camera position and action, medium quality, such as colour or graininess), express what we call a CD-node. An instantiated CD-node can be either based on natural language, such as the CD-node describing the action of a character in FIG. 2, or features or measurements, such as the two CD-nodes in FIG. 2 that express the camera position and movement). In fact, most CD-nodes based on features and measurements will be predominantly generated automatically and mainly serve as the low-level basis for automatic interpretation and generation of material. The two camera-oriented CD-nodes associated with a fragment in FIG. 2 (in reality there would be more, but for the sake of simplicity we just show one for each) provide information on lens movement, lens state, camera distance, movement, position, angle, and production date, they facilitate the establishment of a reference relation automatically, though we assume that most of reference relations will be generated in a semi-automatic way, which is demonstrated by the action-based CD-node, which is directly related to the data via a reference relation. As can be seen, both references are clustered by a unify-relation, which allow to express that the reference of the “station scene” to the “Odessa step scene” is based on cinematographic as well as action parameter.

CA-nodes, finally, are socially determined small reticular semantic structures that allow the interpretation or combination of D-nodes to establish meaning, such as episodes in a narrative, or the identification of metaphors or analogies. In fact, the “unity-relation, that provides the reference cluster, can be understood as CA-node. Another example in FIG. 2 is the description of a particular film technique being attached to the ‘Fragment Odessa Steps’ node. This description can be either Eisenstein’s text or a formalized representation in the form of a feature grammar (see the article by Nack et al. in this issue [44]).

### 4.1.3 The demonstrator – presentational behaviour

The required information network, facilitating not only the representation of ideas and arguments within a work, on the level of higher semantic structures down to the precise unit of articulation, but also the discourse about various aspects of it from different viewpoints. Having established this network it needs to be accessible. For this we designed the demonstrator, built as a Director mock-up. The demonstrator expresses some of the envisioned GUI behaviour, allowing us to get a better feel for the accessibility of such an information space. It must be stressed, though, that the demonstrator had no database, presentation generation or server technology incorporated, and that the behaviour as exemplified in FIG. 3 and FIG. 4, both on the next page, was achieved through LINGO code.

The approach taken for supporting a novice or non-expert to access the complex network structure was based on the assumption that there are various entrance points into the network. These points can be determined based on the taxonomy of roles originally assigned to D-nodes. We are aware of the fact that nodes can play different roles and acquire different meanings depending on the context of their use. However, at the moment of creation each node has a particular purpose. Depending on the ranking within the taxonomy, which is established by the needs of the domain, certain roles are more important than others. For example, when Eisenstein defines a montage technique then this definition is more important than a related example, or even an associated comment. Here, definitions would provide potential entry points. Another option is to define information regions, for example an area of the information space that is predominantly concerned with montage, in which particular entry points are specifically set. These entry points can again be linked to various levels of expertise resulting in appropriate start points.

Once the entrance point is chosen by the user, the system will display the surrounding of the node. The system always tries to present information units of all available media types, as demonstrated in FIG.3, where the images represent videos. The importance of an object is emphasized through its size and position within the presentation area. The importance can be determined on aspects such as traversal-valency or significane (relation between link type and context of investigation) of a relation between nodes. FIG 3 shows the full text of the definition for intellectual montage, whereas the related textual examples are only visualised as an active area, encouraging further investigation. A visual example, on the other hand, is displayed more prominently just above it. For a similar approach on three-dimensional graphic representations of large knowledge spaces see [24]. In some respect we also overlap here with the work provided by Miles [37] and Mancini [35], which are both concerned with topographical issues.

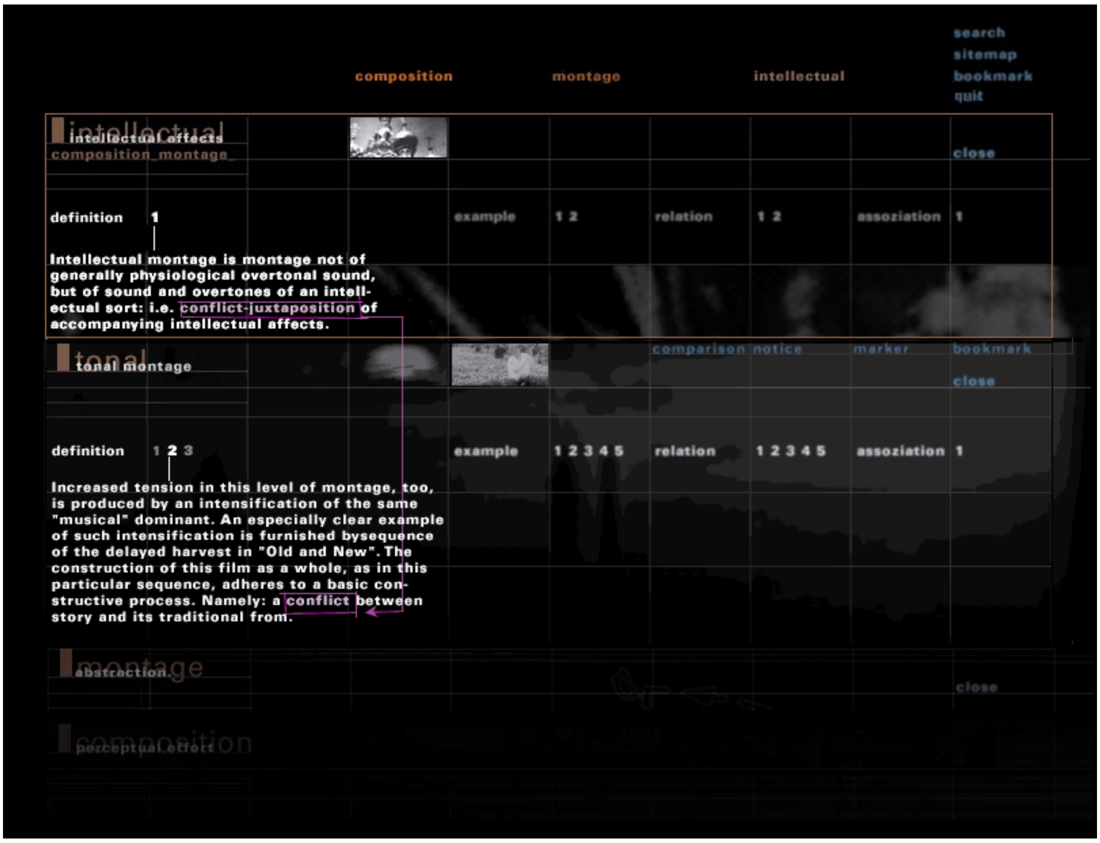


FIG 3: Dynamic Interface representing the various ways of linking material


FIG 4: Dynamic Interface representing the importance of a single medium based on size

The user indicates an information need, i.e. the explanation of interest in an information unit, through presenting, pointing at, or otherwise indicating it. For example, the user might activate the highlighted "conflict" area in the definition of tonal montage (see FIG. 3). This part is highlighted because it indicates the link to another information unit. Due to the representation structure defined above, we are able to link into data and thus establish precise context. This concept is now transferred into a presentation where

- the change of context (the link in the network relates two different clusters, i.e. tonal and intellectual montage) is introduced by a transition of the former presentation to the bottom of the screen, being grouped within a frame, and a fade in of the new information units related to the chosen node, also combined in a frame;
- the two related parts in the two most important information units, the definitions, are not only highlighted, but also connected via a visible line, guiding the user's attention to the interesting areas.

FIG. 4 is an example where the change of topic was initiated not only by shifting to another topic but also by putting an emphasis on a different media. The importance of the visual media is documented by the larger space occupied in the presentation area. The result of such a media change is that the system will now try to collect more media items of that type for the ongoing presentation of the information quest. Moreover, since the system is provided with knowledge of low-level denotative information about visual media items, it can highlight regions or objects and generate visual links in the same way as described above. However, a link must not be drawn but can also be displayed in the form of an overlay, where the related abstract concept is add to the video in the form of a caption whenever needed (using the temporal information from the setting definition).

Thus, the items on the screen represent a fraction of the items potentially available. As the user browses the generated interface, the system uses the distilled information about the user's preferences on media, topics, argument and presentation complexity, and so forth, to create the following visualisation.

The dynamic concept of the browser (investigation of node surroundings) allows the representation of the relevance of information based on spatial, textual and temporal properties of the different information units. The relevant information for such representations is retrieved from content description nodes, as well as from the syntactical properties of a relation or conceptual annotation nodes. Here we are following the general ideas on adaptive hypermedia [15, 20], though our approach seems to be more flexible since it allows the incorporation of various points of view on a particular topic over time.

Thus, users can investigate an unknown space provided with the most relevant material and its annotations for the actual need. This facilitates a progressive experience that completes the understanding of complex concepts by procedural and participatory means, e.g., an interactive investigation in a navigable encyclopaedic space that provides access to the full media items. Such an experience can yield an understanding of a concept more primal and powerful than any appeal through normal text in a linear logical form.

However, we realised while presenting the demonstrator in a number of design competitions [57, 58] and at various public presentations in the institute that the dynamic concept of the interface was too difficult for most of the audiences. The idea of moving information units, even though clustered in frames, that change size and might even disappear after some time took most users by surprise and only younger users, in the age range of 12 – 25 years had no problems in adjusting swiftly. Moreover, it took also time to get accustomed to the various navigational codes, such as the colour codes for link context. On the other hand, just by playing with the demonstrator, users had the feeling that they had learned a lot about Eisenstein's work and they wished they had more time to explore the information space even further.

It became apparent that the envisioned system requires a large amount of development work. Thus we will use the next section to outline the environment required for the development and maintenance of the illustrated information network.

## 4.2 An architecture to establish and maintain semiosis-aware hypermedia

The presented Eisenstein demonstrator is only an example of how hypermedia can be applied in the context of theory, history, and anthropology of film. Moreover, we are aware that other domains, such as chemistry or architecture, might require different representations on a semantic level. However, it is obvious that the development and, even more important, the maintenance of a complex and dynamic hypermedia knowledge space requires a flexible working environment. To our knowledge no such environment exists at the time of writing, because as yet, no large knowledge space as the one described, exists. In this section we therefore focus on the environmental necessities, which will allow us to work towards large hypermedia knowledge spaces. We wish to point out that not everything described below is already implemented in our system [45].

An environment that allows the incorporation of different media units on a semiotic level for intelligent information cognition has to serve three functional levels:

- It is a digital data repository for a variety of primary data (text, audio, video, 3D animation, 2D image, 3D image, graphic) and the expert annotations related to one or more of these primary data.
- Metadata management tools are needed to organize the stored documents to facilitate the complex knowledge-intensive tasks that domain experts, design experts and users want to perform.
- An extra service layer that reflects the domain objects and related user tasks is needed. These tasks include indexing, annotation, retrieval and the presentation of retrieval results for specific publication purposes.

The system architecture sketched in FIG. 5 describes the main modules of a semiotics-aware hypermedia system and the way in which they interact with each other.

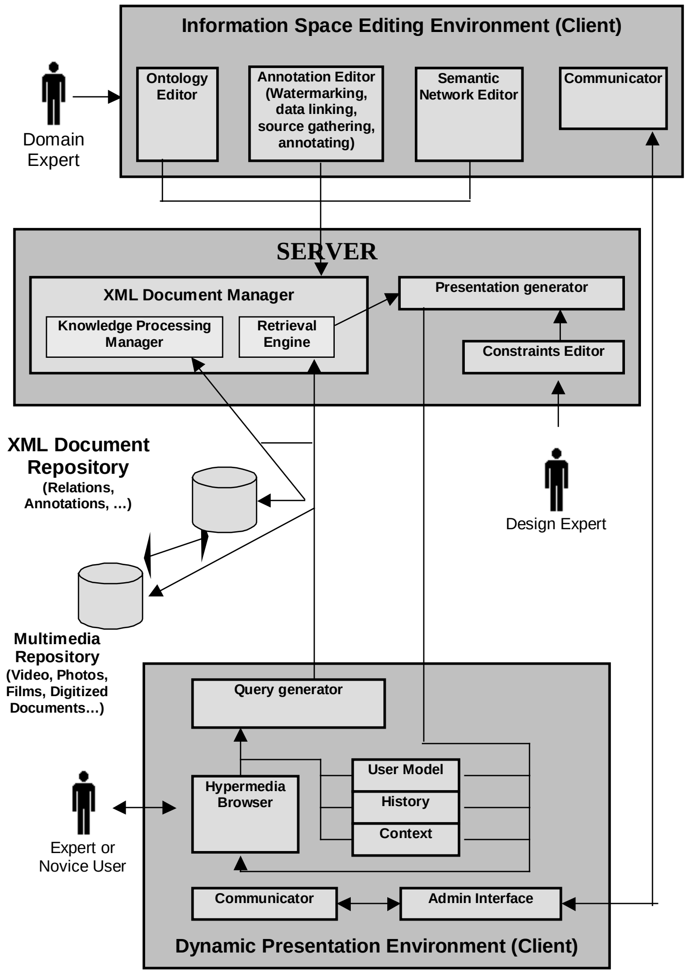


FIG 5: Proposed architecture for a dynamic semiotic-aware hypermedia system

### 4.2.1 Information Space Editing Environment

The **Information Space Editing Environment** (see top of FIG. 5) serves the specific needs of the domain expert. The suggested approach is that all indexing and classifying instruments will be controlled by the annotation interface to guide and support experts. The definitions of schemata for denotative and connotative descriptions and the required taxonomy of relations will be left to the domain experts. However, the encoding processes of instantiations and the spatio-temporal identification marks for audio-visual data, based on a linking mechanism using time-codes and region-codes, into the internal XML representation, should be generic. This allows the architecture to be applied to any domain. The advantage of the generic approach is that experts can concentrate on their research tasks without being concerned about storage organisation or general presentation. The particular tools necessary are as follows:

The **Ontology editor** allows the definition of concepts and relations in the form of task-specific controlled vocabulary/subject indexing schemata for the in-depth semantic indexing of various media. Concepts and relations should be described by standardised structures [26].

The **Annotation editor** permits the splitting of the media content into small information units, their description on a denotative level, and the establishment of their conceptual relationships (connotative level). The description process is controlled by the defined ontology and follows a strata-oriented approach [1], which allows a fine-grained sequence and region description of media content. By placing an annotation or a relation the creator leaves a mark in the network, since every placed item has an author. Examples of such environments are provided by [4, 42, 43].

The **Semantic network editor** provides a means of defining rhetorical structures for information units. They are also ontology-controlled and are used during presentation in combination with the user profile and the presentation plans for the

delivery of content.
The **Information Space Editing Environment** allows experts to control the quality of the content, given their exclusive access to the tools, a necessity based on the idea of 'perspective making' [9]. The availability of typed concepts and relations in combination with personal identification of contributions allows decisions to be made about the impact, evidence, and consistency of work as well as automatically providing classifications such as schools of thought or research positions. However, permitting non-experts to contribute to the growing hypermedia space, a communicator tools suite is required, which is mainly a state-of-the-art collaboration environment allowing non-experts to exchange any sort of data, such as annotated screen shots, text and audiovisual material with the experts. If the submitted material or the suggested relation is incorporated into the network, then the external user also leaves a mark.

### 4.2.2 Document management

The repository modules that store the XML-based meta-data (content description nodes, conceptual annotation nodes and the relations among them) and the multimedia data (data -nodes) can be implemented using federated database technology. Access to them is governed by the **XML Document Manager** (see the Server box in FIG. 5), which also supports the search for specific information units by the **Retrieval Engine** that uses the relevant document structures, relations and links.
The **Presentation generator** is a constraint-based planning system that uses information about the user and the definitions provided by a design specialist or the content space owner [45]. The Presentation generator analyses the retrieved material based on the user model, redesigns the new presentation according to design issues such as graphic direction, scale, volume, depth, style (i.e. physical manipulation of the material for better integration into the presentation), temporal synchronisation, etc., and provides a format, that a hypermedia browser can interpret (e.g. SMIL [54] or MPEG-4 [29]). Simultaneously, the Presentation generator also updates the user model (e.g. user preferences), browsing history and the current context setting on the client side.

### 4.2.3 Dynamic Presentation Environment

We already outlined this environment in the discussion of the Eisenstein demonstrator. However, since then parts of the described mechanisms found their implementation in our Cuypers project [45], where we use the mapping of high-level conceptual structures with low-level feature descriptions as an essential mechanism to enhance the automatic generation of dynamic style-oriented multimedia environments for the domain of musea. A detailed description of the underlying mechanisms can be found in our other article in this issue [46].

## 5 CONCLUSION

In this article we have argued that the linguistic-centred view within hypermedia systems needs extension through a semiotic-based approach if real analytic parity among media is to be achieved. We discussed the problems of visual signification for images and video in dynamic systems, where users can access visual material in an interactive fashion. We described how semiotics can help to overcome such problems by allowing descriptions of the material both on a denotative and connotative level and proposed an architecture for a dynamic semiotic-aware hypermedia system.
While our research is concerned with the automatic generation of hypermedia presentations, the facilities described above are far from being in place. In addition, the approach described in this article is but a small step towards a real semiotic-oriented hypermedia system. To achieve the goal of semiotic-aware presentation systems, it will be necessary to do the following:

- improve existing search techniques on category and media features,
- develop better presentation techniques that incorporate presentation-independent knowledge (e.g. user profiles generated over time during the browsing session, graphic design rules, and domain knowledge as described by the expert);
- elaborate on automated generation processes at different abstract levels, i.e. communication strategies (hypermedia design patterns [6, 63] qualitative, quantitative and multidimensional constraints [50, 51], and media integration (e.g. assigning filters to media objects to vary their style for closer integration with the rest of the presentation [61, 64]).

For a smooth integration of media with the user interface and for accessing regions and objects within media data one can combine MPEG-4 and SMIL technology. This approach provides the required flexibility for dynamic user-centred presentations, which current state of the art Web technology cannot. This Web technology includes presentation languages, such as SMIL (integration of media style [66]), SVG (with CSS for graphics [67]) and XHTML (with CSS for formatted text [68]), or transformational methods, such as XSLT (document transformation [69]) and CSS (control of style appearance [65]). The research challenge will be to incorporate and improve these by extending the relevant DTDs or schemata. A detailed description of such problems is provided by Ossenbruggen et al. [46].
We would like to stress the fact that digitised audio-visual material provides an important opportunity for hypermedia systems – both for authoring as well as for presentations.
We believe that our view on semiotic-based hypermedia environments provides an essential foundation for adaptive and adaptable hypermedia, facilitating environments in which complex domain information can be studied, discussed,

commented, published and demonstrated – thus preserving and developing information artefacts in an intelligent communal way.

## 6 ACKNOWLEDGMENTS

The authors wish to thank Helma Töpper who implemented the Eisenstein demonstrator in Director for the user tests. The development of the demonstrator was funded by GMD-IPSI in Germany. Moreover, the authors would like to thank Werner Kriechbaum from IBM Böblingen, Germany, for insightful discussion and helpful comments. Finally, we'd also like to acknowledge Clara Mancini and Simon Buckingham Shum from the Knowledge Media Institute at the Open University, UK, for useful discussions during the development of this work. Finally, we wish to thank the anonymous reviewers of NRHM for their detailed and valuable comments. Part of this research was funded under the Dutch NWO research projects Dynamo and ToKeN 2000.

## 7 REFERENCES

[1] Aguierre Smith, T. G., & Davenport, G. (1992). The Stratification System. A Design Environment for Random Access Video. In ACM workshop on Networking and Operating System Support for Digital Audio and Video. San Diego, California

[2] Andrew, D. (1984). *Concepts in Film Theory*. Oxford: Oxford University Press.

[3] Arnheim, R. (1956). Art and Visual Perception: A Psychology of the creative eye. London: Faber & Faber.

[4] Auffret G., Carrive J., Chevet O., Dechilly Th. Ronfard R. & Bachimont B. (1999). Audiovisual-based Hypermedia Authoring: using structured representations for efficient access to AV documents. Proceedings of the 10th ACM conference on Hypertext and Hypermedia, pp. 169-178, February 21-25, Darmstadt, Germany.

[5] Barthes, R. (1977). *Image, Music, Text*. Fontana Press, London.

[6] Bernstein, M. (1998). *Patterns of Hypertext*. Proceedings of the 9th ACM conference on Hypertext and Hypermedia, pp. 21 - 29, June 20-24, Pittsburgh, USA.

[7] Bloch, G. R. (1986) *Elements d'une Machine de Montage Pour l'Audio-Visuel*. Ph.D., Ecole Nationale Supérieure Des Télécommunications.

[8] Boissière, G. (1998). *Automatic Creaton of Hypervideo News Libraries for the World Wide Web*. Proceedings of the 9th ACM conference on Hypertext and Hypermedia, pp. 279 -280, June 20-24, Pittsburgh, USA.

[9] Boland, R.J.J. and Tenkasi, R.V. (1995). *Perspective Making and Perspektive Taking in Communeties of Knowing*. Organization Science, 6 (4), 350–372.

[10] Bordwell, D. (1989). Making Meaning - Inference and Rhetoric in the Interpretation of Cinema. Cambridge, Massachusetts: Harvard University Press.

[11] Brachman, R.J. & Levesque, H.J. (1983), *Readings in Knowledge Representation.* San Mateo, California: Morgan Kaufmann Publishers.

[12] Burch, N. (1981). *Theory of Film Practice*. Princeton, New Jersey: Princeton University Press.

[13] Buckingham Shum S., DomingueJ., & Motta E. (2000). Scholarly Discourse as Computable Structure, Technical Report KMI-TR-93. The Knowledge Media Institute, Open University, UK, April 25, 2000.

[14] Buckingham Shum, S. & Selvin, A. (2000). Structuring Discourse for Collective Interpretation. Electronic proceedings of Open Conference on Collective Cognition and Memory Practices, September 19-20, 2000. http://www.limsi.fr/WkG/PCD2000/indexeng.html

[15] Brusilovsky, P., Kobsa, A., and Vassileva J.(ed.) (1998).*Adaptive Hypertext and Hypermedia*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

[16] Carter, L. M. (1997). *Arguments in Hypertext: A Rhetorical Approach*. Proceedings of the 11th ACM conference on Hypertext and Hypermedia, pp. 87 - 91, May 30 – June 4, San Antonio, Texas, USA.

[17] Chiu, P. Foote, J., Girgensohn, A., and Boreczky J. (2000). *Automatically linking Multimedia Meeting Documents by Image Matching*. Proceedings of the 11th ACM conference on Hypertext and Hypermedia, pp. 244 - 245, May 30 – June 4, San Antonio, Texas, USA.

[18] Collier G. H. (1987). "Thoth-II: Hypertext with Explicit Semantics" in *Proceedings of ACM Hypertext '87*, pp. 269-289, Chapel Hill, North Carolina, ACM Press, November 13-15, 1987.

[19] Davis, M. (1995) *Media Streams: Representing Video for Retrieval and Repurposing*. Ph.D. Thesis, MIT.

[20] De Bra, P., and Calvi. L. (1998). *AHA: A Generic Addaptive Hypermedia System*. Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia, Pitsburgh, June 1998, pp. 5 – 17.

[21] Eco, U. (1976). *Articulations of the Cinematic Code*. In B. Nichols (Eds.), *Movies and Methods* (pp. 590 - 607). Berkeley: University of California Press.

[22] Eisenstein, S. M. (1988). *Selected Works: Writings 1922 - 1934*. London: BFI Publishing.

[23] Eisenstein, S. M. (1991). *Selected Works: Towards a Theory of Montage*. London: BFI Publishing.

[24] Fairchild, K. M., Poltrock, S. E., and Furnas, G. W. (1988). SemNet: Three-Dimensional Graphic Representations of Large Knowledge Bases. Cognitive Science and its Applications for Human-Computer Interaction, Raymonde Guindon (ed.), pp. 201 – 233. Laaawrence Erlbaum Associates, Hillsdale, New Jersey.

[25] Gregory, J. R. (1961) *Some Psychological Aspects of Motion Picture Montage*. Ph.D. Thesis, University of Illinois.

[26] Grosso, W.E, Eriksson, H., Fergerson, R.W., Gennari, J. H., Tu, S.W., and Musen, M.A. (1999). *Knowledge Modeling at the Millennium (The Design and Evolution of Protege-2000).* Stanford Medical Informatics (SMI), SMI Report Number: SMI-1999-0801. Stanford, USA.

[27] Halasz, F. G. (1988). "Reflection On Notecards: Seven Issues For The Next Generation Of Hypermedia Systems" in *Communications of the ACM*, July 1988, 31 (7).

[28] Ingarham, B. D. (2001). *Scholary Rhetoric in Digital Media*. Journal of Interactive Media in Education. http://www-jime.open.ac.uk/00/ingraham/ingraham-t.html.

[29] ISO MPEG-4 (2000). MPEG-4 Overview - (V.15 - Beijing Version), ISO/IEC JTC1/SC29/WG11 N3536, Beijing - July 2000

[30] Jakobson, R., & Halle, M. (1980). *Fundamentals of Language*. The Hague: Mouton Publishers.

[31] Kolb, D. (1997). *Scholarly Hypertext: Self-Represented Complexity*. Proceedings of the 8th ACM conference on Hypertext and Hypermedia, pp. 29 - 37, April 6-11, Southampton, UK.

[32] Kuleshov, L. (1974). *Kuleshov on Film - Writing of Lev Kuleshov*. Berkeley: University of California Press.

[33] Landow, G. P. (1994). *The Rhetoric of Hypermedia: Some Rules for Authors*. Hypermedia and Literary Studies. Eds. Paul Delany and George P. Landow, pp. 81 – 103, Cambridge: The MIT Press.

[34] Liestøl, G. (1994). *Aesthetic and Rhetorical Aspects of Linking Video in Hypermedia*. Proceedings of the ACM European Conference Hypermedia Tchnology 94 (*ECHT '94), pp. 217 - 223*.

[35] Mancini, C. (2000). *From Cinematographic to Hypertext Narrative*. Proceedings of the 11th ACM conference on Hypertext and Hypermedia, pp. 236 -237, May 30 – June 4, San Antonio, Texas, USA.

[36] Metz, C. (1974). *Film Language: A Semiotic Of The Cinema*. New York: Oxford University Press.

[37] Miles A. (2000). *Hypertext in the Dark: cinematic narration with links / Hypertext Syntagmas: Cinematic Narration with Links*. Journal of Digital Information, Vol. 1, No. 7, December 2000 http://jodi.ecs.soton.ac.uk/Articles/v01/i07/Miles/

[38] Miyasato, T. (2000). *Creation of Interactive Media Content by the Reuse of Images*. Proceedings of the 11th ACM conference on Hypertext and Hypermedia, pp. 246 - 247, May 30 – June 4, San Antonio, Texas, USA.

[39] Monaco, J. (1981). *How To Read A Film*. New York: Oxford University Press.

[40] Moulthrop, S. (1991). *Beyond the Electronic Book: A Critique of Hypertext Rhetoric*. Proceedings of the 4th ACM conference on Hypertext, pp. 291 - 298, December 15-18, San Antonio, USA.

[41] Nack, F. (1996) "AUTEUR: The Application of Video Semantics and Theme Representation in Automated Video Editing," Ph.D., Lancaster University, 1996.

[42] Nack, F. and Lindley, C. (2000) Production and maintenance environments for interactive audio-visual stories. In ACM MM 2000 WS proceedings - Bridging the Gap: Bringing Together New Media Artists and Multimedia Technologists, pp. 21 –24, October 31, 2000 Los Angeles, CA.

[43] Nack, F. & Putz, W. (2001). Designing Annotation Before It's Needed. In Proceedings of the 9th ACM International Conference on Multimedia, pp. 251 - 260, Ottawa, Canada, Sept. 30 - Oct. 5, 2001.

[44] Nack, F., Windhouwer, M., Hardman, L., Pauwels, E., and Huijberts, M. (2001). *The Role of High-level and Low-level Features in Style-based Retrieval and Generation of Multimedia Presentations*. New Review of Mypermedia and Multimedia (NRHM) 2001.

[45] Ossenbruggen, J. v, Cornelissen F., Geurts J., Rutledge L., & Hardman L. (2001) Towards second and third generation Web-based multimedia In: The Tenth International World Wide Web Conference, May 1-5, 2001, Hong Kong

[46] Ossenbruggen, J. v, Hardman L. & Rutledge L. (2002) Hypermedia and the Semantic Web: A Research Agenda. To be published in Journal of Digital Information (Jodi), 2002.

[47] Parkes, A. P. (1989). *An Artificial Intelligence Approach to the Conceptual Description of Videodisc Images*. Ph.D. Thesis, Lancaster University.

[48] Pasolini, P. P. (1976). *The cinema of poetry*. In B. Nichols (Eds.), *Movies and Methods* (pp. 542 - 558). Berkeley: University of California Press.

[49] Peirce, C. S. (1960). *The Collected Papers of Charles Sanders Peirce - 1 Principles of Philosophy and 2 Elements of Logic, Edited by Charles Hartshone and Paul Weiss.* Cambridge, MA: The Belknap Press of Harvard University Press.

[50] Rutledge, L., Davis, J., van Ossenbruggen, J., & Hardman L. (2000). Inter-dimensional Hypermedia Communicative Devices for Rhetorical Structure In: Proceedings of International Conference on Multimedia Modeling 2000 (MMM00), November 13-15, 2000, Nagano, Japan

[51] Rutledge, L., Davis, J., van Ossenbruggen, J., & Hardman L. (2000). Inter-dimensional Hypermedia Communicative Devices for Rhetorical Structure In: Proceedings of International Conference on Multimedia Modeling 2000 (MMM00), November 13-15, 2000, Nagano, Japan

[52] Rutledge L., van Ossenbruggen J., Hardman L., & Bulterman D.C.A. (1999). Mix'n'Match: Exchangeable Modules of Hypermedia Style. Proceedings of the 10th ACM conference on Hypertext and Hypermedia, pp. 179-188, February 21-25, Darmstadt, Germany.

[53] Smith, J. Mc., Stotts, D., and Kum, S-U.(2000). *An Orthogonal Txonomy for Hyperlinking Anchor Generaton in Video streams Using OvalTine.* Proceedings of the 11th ACM conference on Hypertext and Hypermedia, pp. 11 - 18, May 30 – June 4, San Antonio, Texas, USA.

[54] SMIL (2001). http://www.w3.org/TR/2001/REC-smil20-20010807/

[55] Saussure, F. d. (1966). *Course in General Linguistics - edited by Charles Balley, Albert Sechehaye and Albert Riedlinger*. New York: McGraw-Hill.

[56] Sawhney, N., Balcom, D., and Smith, I. (1996). *HyperCafe: Narrative and Aesthetic Properties of Hypervideo*. Proceedings of the 7th ACM conference on Hypertext, pp. 1 - 10, March 16 - 20, Washington DC, USA.

[57] Sowa, J. F. (1984). *Conceptual Structures: Information Processing in Mind and Machine*. Reading, MA: Addison-Wesley Publishing Company.

[58] Toepper, H. (1999) Sergej Eisenstein – a documentation about his work and life. The Europrix '99 students' Contest, Europe's Top Talent Inovation and Content Creation, pp. 60-63. EuroPrix Secretariat (also on CD-Roma) http://www.europrix.org/

[59] Toepper, H. (2000) Sergej Eisenstein – a documentation about his work and life. 1. Workshop on Digital Storytelling, Darmstadt, Germany, 15-16/6/2000, pp. 59 – 72. http://www.zgdv.de/distel2000/

[60] Tosca, S. P. (2000). A Pragmatics of Links. Proceedings of the 11th ACM conference on Hypertext and Hypermedia, pp. 77 - 84, May 30 – June 4, San Antonio, Texas, USA. Also available in Journal of Digital Information, Vol. 1, No. 6, June 2000, http://jodi.ecs.soton.ac.uk/Articles/v01/i06/Pajares/

[61] Vries, de A.P., Windhouwer,M.A., Apers,P.M.G., & Kersten, M.L. (2000). Information Access in Multimedia Databases based on Feature Models. CWI Internal Report, January 2000

[62] Walker, J. (1999). *Piercing together and tearing apart: finding the story in afternoon*. Proceedings of the 10th ACM conference on Hypertext and Hypermedia, pp. 111-117, February 21-25, Darmstadt, Germany.

[63] Weitzmann, L. and Wittenburg, K. Automatic presentation of multimedia documents using relational grammars. In: Proceedings of the second ACM international conference on Multimedia '94 San Francisco pp. 443-451, October 15 - 20, 1994

[64] Windhouwer, M.A., Schmidt,R.A. & Kersten, M.L. (1999).. Acoi: A System for Indexing Multimedia Objects. In International Workshop on Information Integration and Web-based Applications & Services, Yogyakarta, Indonesia, November 1999.

[65] W3C – CSS (1999). Cascading Style Sheets, level 1, W3C Recommendation 17 Dec 1996, revised 11 Jan 1999. http://www.w3.org/TR/REC-CSS1

[66] W3C – SMIL (2001). Synchronized Multimedia Integration Language (SMIL 2.0). W3C Recommendation 07 August 2001. http://www.w3.org/TR/smil20/

[67] W3C – SVG (2001). *Scalable Vector Graphics (SVG)*. W3C Recommendation 4. September 2001. http://www.w3.org/TR/SVG/

[68] W3C – XHTML (2001). The Extensible HyperText Markup Language, A Reformulation of HTML 4 in XML 1.0, W3C Recommendation 26 January, 2000. http://www.w3.org/TR/xhtml1/

[69] W3C – XSLT (2001). XSL Transformations (XSLT) Version 1.0. W3C Recommendation 16 November 1999. http://www.w3.org/TR/xslt

[70] Zellweger, P. T., Chang, B-W., and Mackinlay, J. D. (1998). *Fluid Links for Informed and Incremental Link Transitions*. Proceedings of the 9th ACM conference on Hypertext and Hypermedia, pp. 50 - 57, June 20-24, Pittsburgh, USA.