

KBM – Video Application

Frank Nack



Outline

- Organisation
- Last lecture
- AUTEUR – automatic generation of slapstick video sequences
- Additional applications

Video – summary

Investigated

- Temporal visual medium
- Effect of time on interaction and creativity
- AI techniques for the representation of time, action, plan

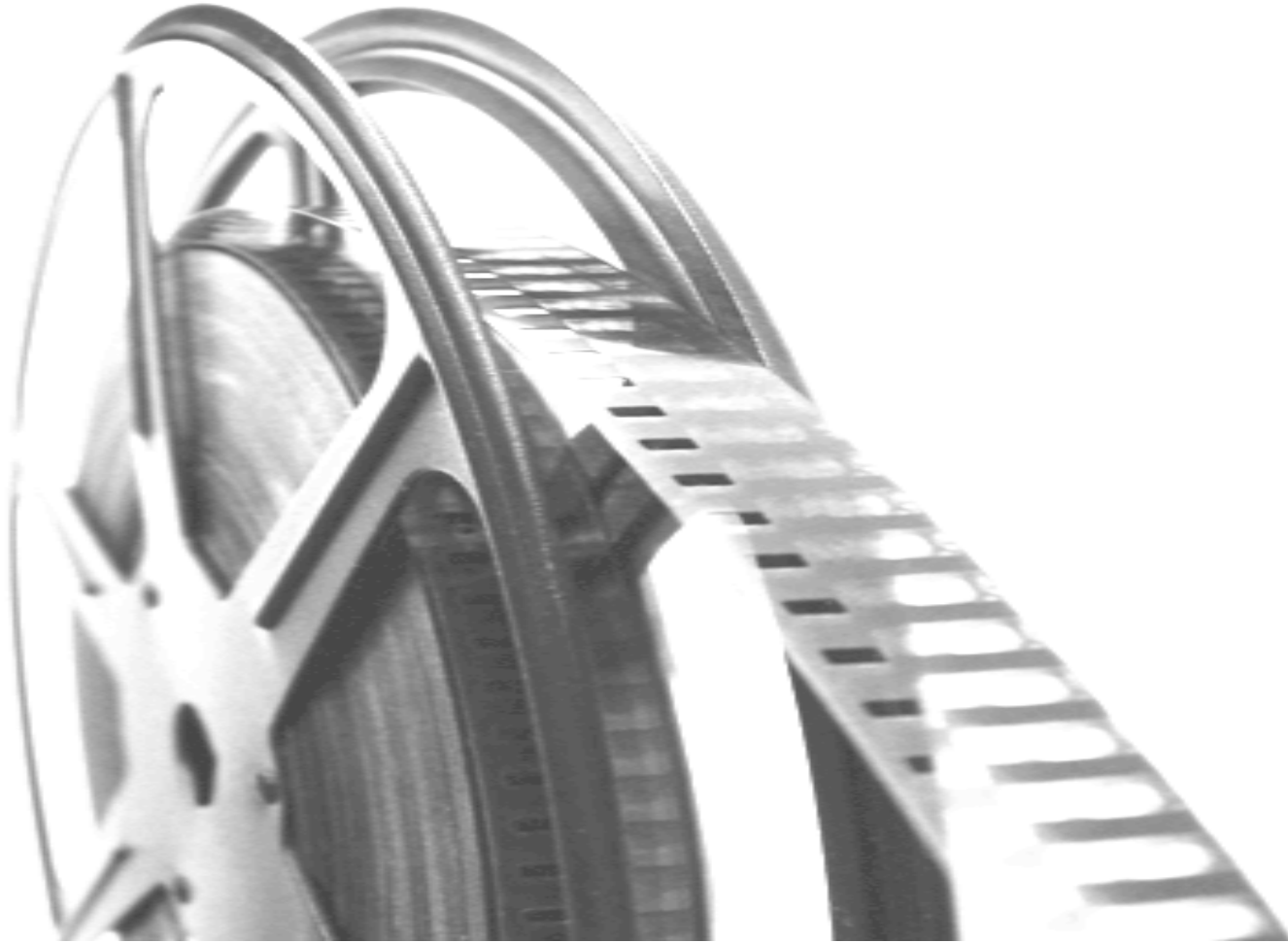
Example

Video, Frames, Scripts, Conceptual Dependencies, Plans, Goals, Story Grammar, Semantic Links

Findings

- Video, though based on common human content and thematic structures, provides its own realities of time and space which are interwoven in the narrative structure.
- A story is a representational system based on two main layers, structure and content, each serving two distinct purposes (form and substance).
- A primarily structure-oriented approach (grammar) to the temporal aspects of video with respect to generation and interaction is not appropriate.
- A planning approach (planner or agent) seems more workable for automatic video generation.

AUTEUR



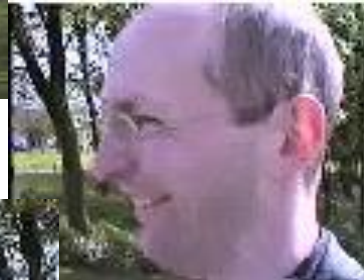
AUTEUR - Aim



- Automatic composition of visual slapstick sequences.
- Use an existing, arbitrary data base.
- Investigate machine creativity.

AUTEUR – Generation Example I

Motivation

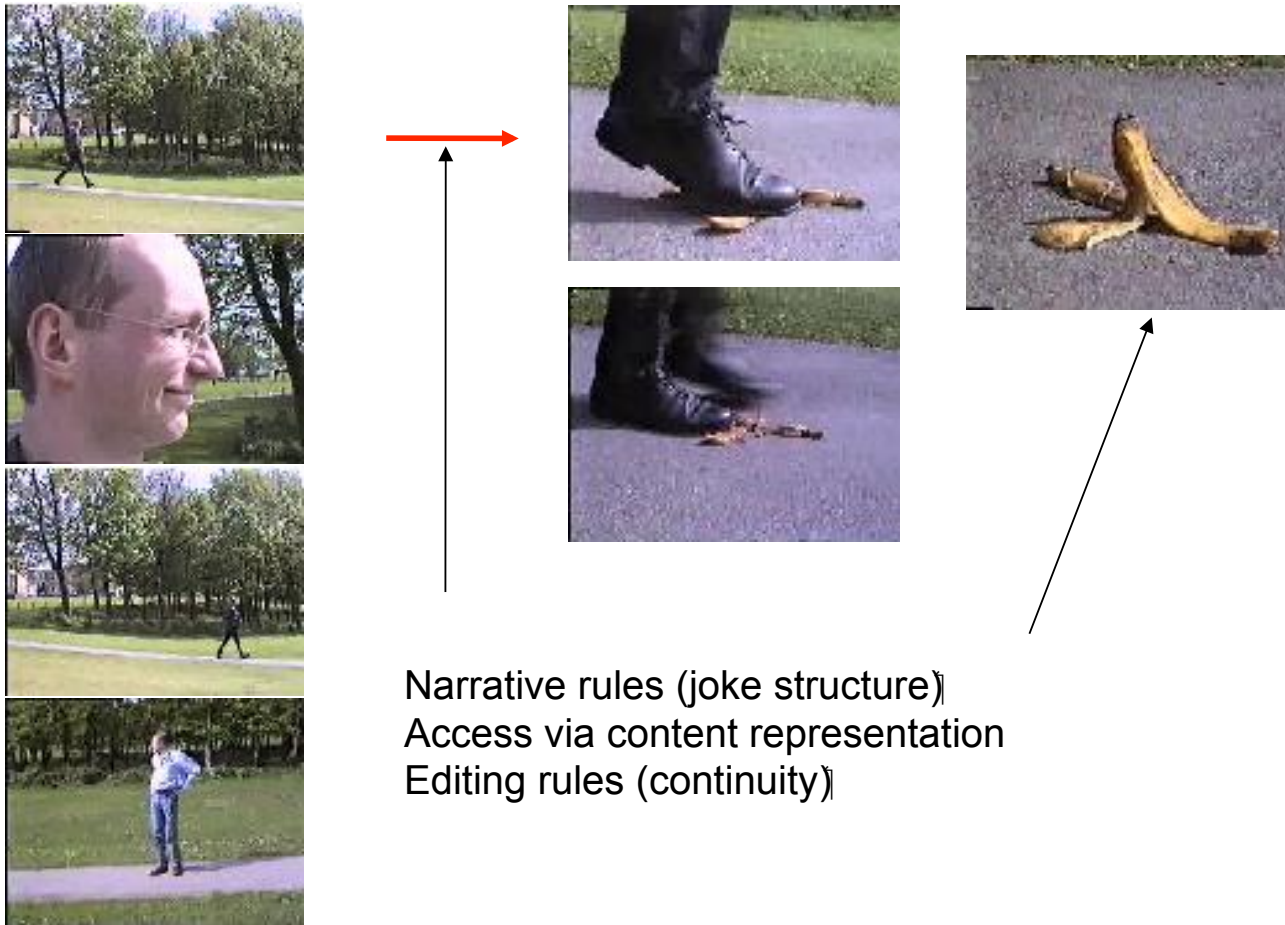


Narrative rules (joke structure)
Access via content representation
Editing rules (continuity)

Narrative rules (continuity)
Access via content representation
Editing rules (continuity)

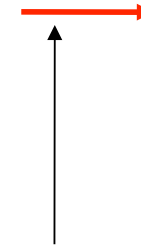
AUTEUR – Generation Example II

Realisation



AUTEUR – Generation Example III

Resolution



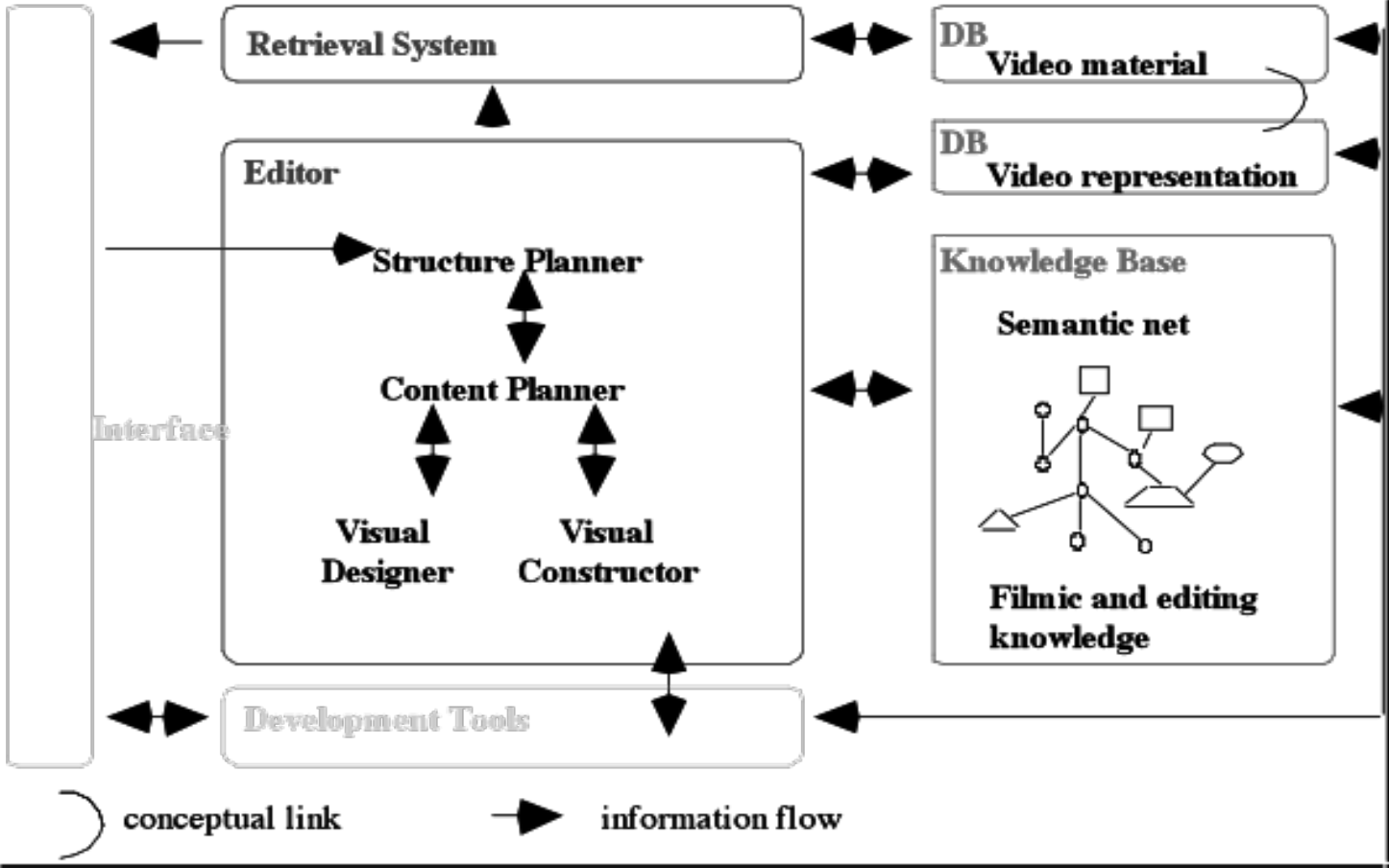
Narrative rules (joke structure)
Access via content representation
Editing rules (continuity)

AUTEUR – Generation Example IV

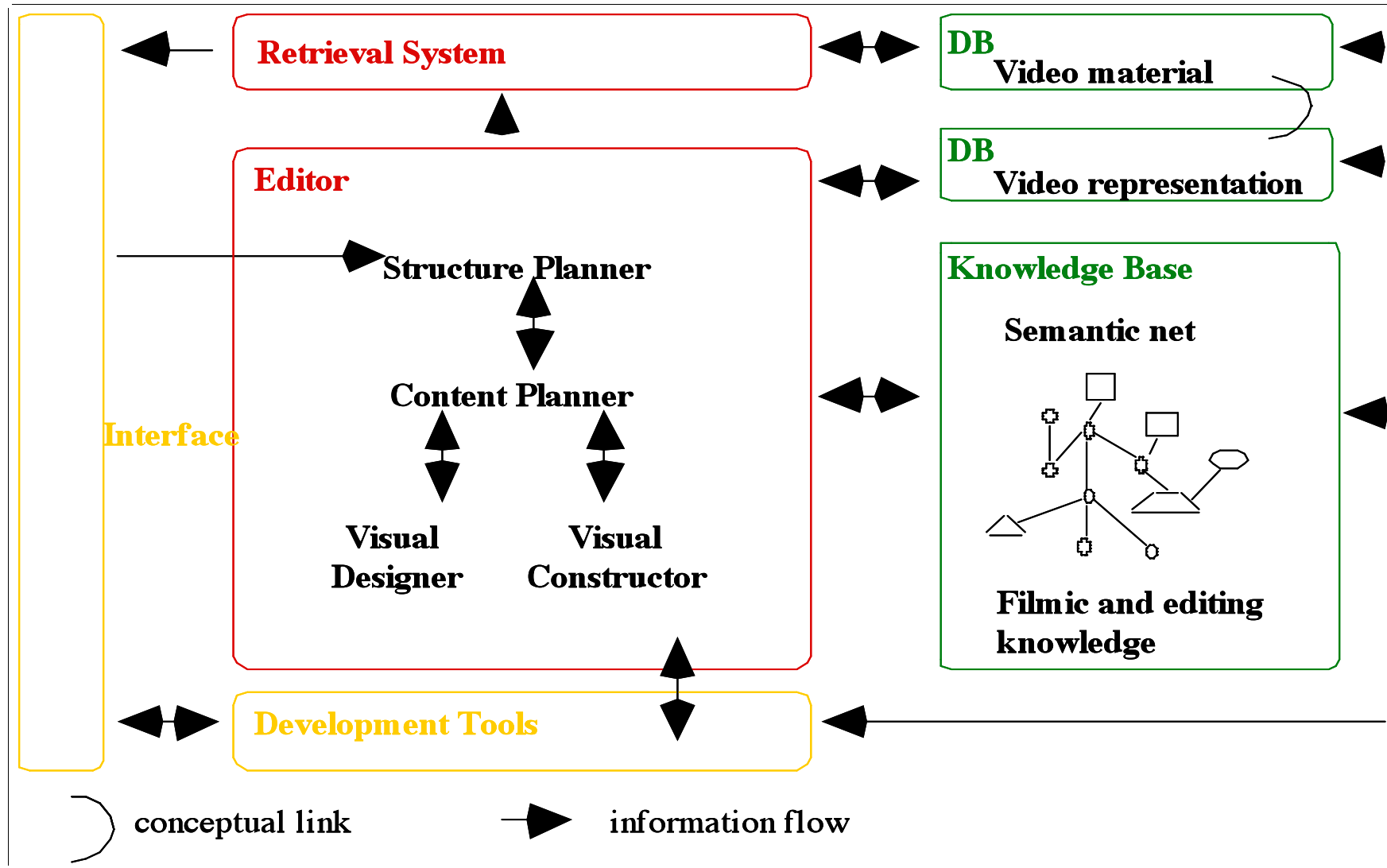
Result



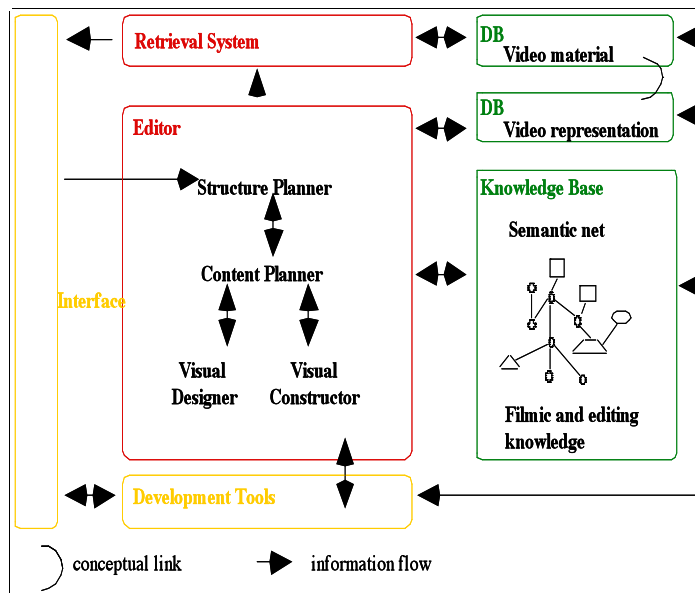
AUTEUR - Architecture



AUTEUR - Architecture

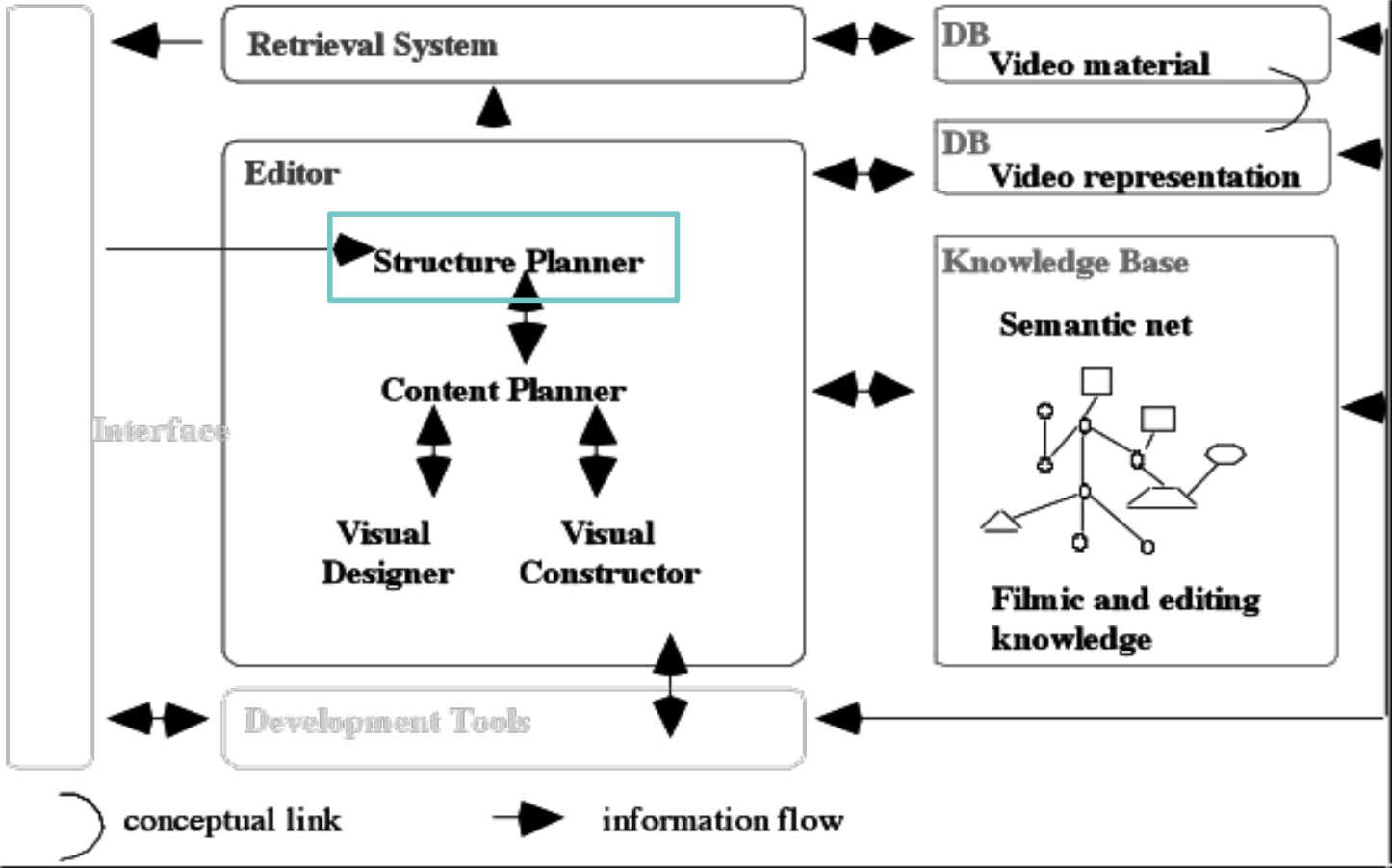


AUTEUR - Architecture

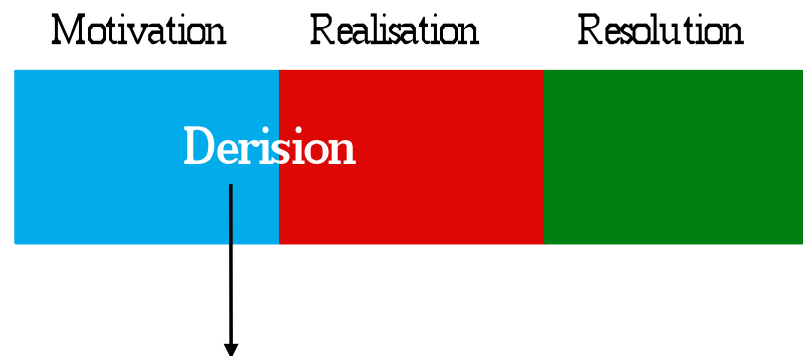


- 145 Conceptual Dependencies
- 26 humour rules
- 37 editing rules
- 85 clips (5 sec – 9 min)

AUTEUR – Structure Planner

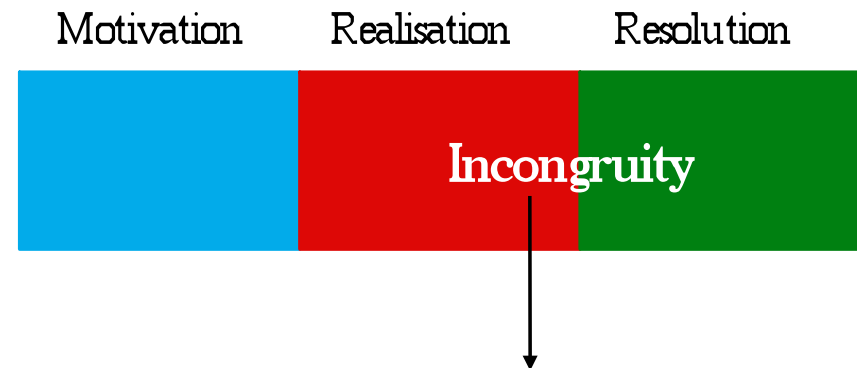


AUTEUR – Plot Model I



The explicit transformation of moods or states of the portrayed character.

The implicit upgrading of the mood of the viewer.

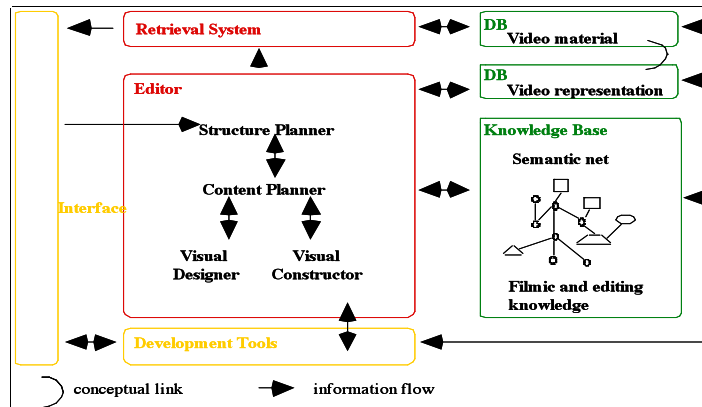


The conflict between the expected and what actually occurs.

AUTEUR – Plot Model II



AUTEUR – Humour Planner



H-Strategy 1

An action forms the most suitable subject for a joke, then an actor, then an object, and finally a location.

H-Strategy 2

If the action portrays an intention [goal], interrupt the action in a way that is unexpected by the character, so that the goal cannot be fulfilled and the character's mood is downgraded or he or she suffers in some way. (Mischief + Schadenfreude)

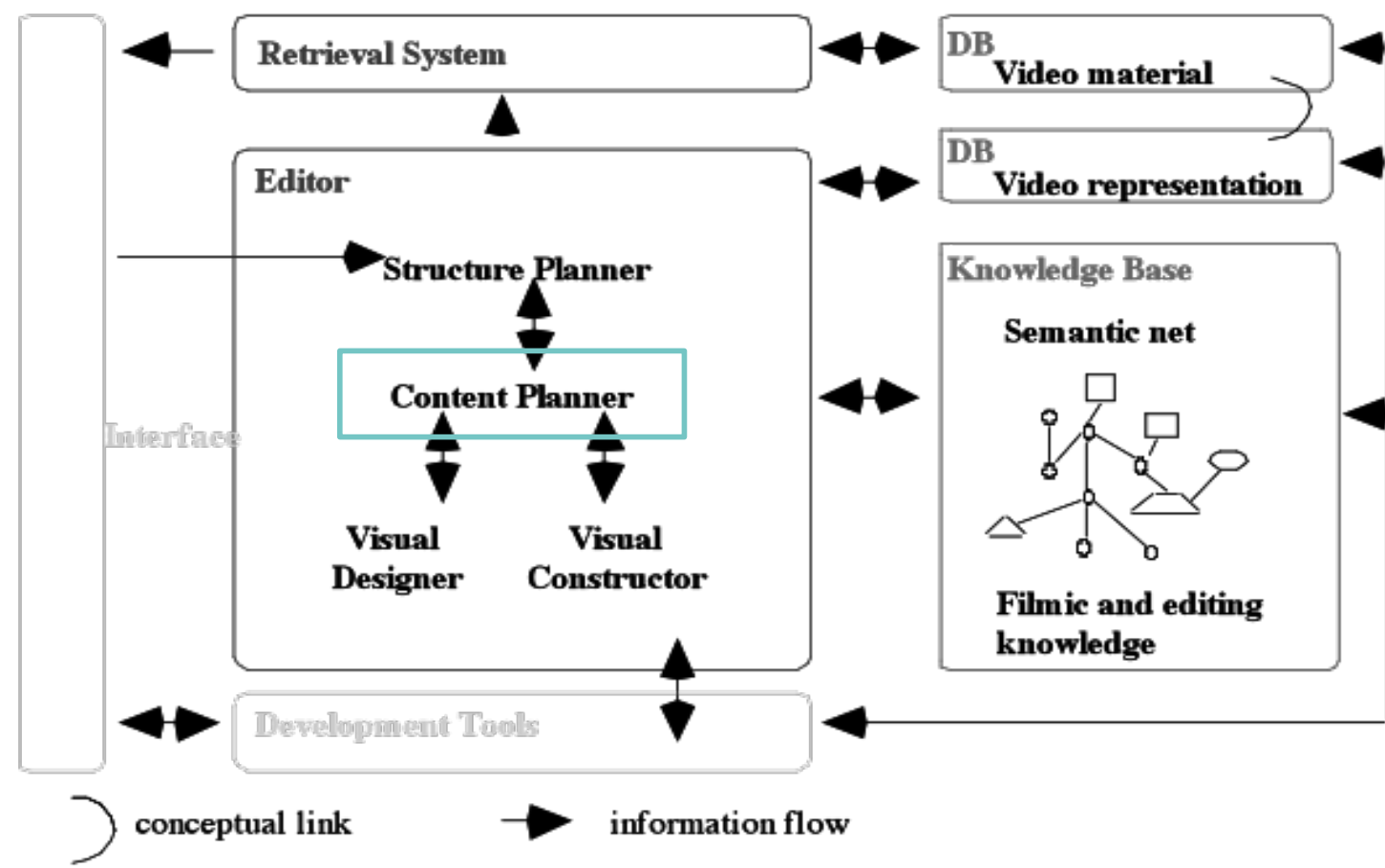
H-Strategy 16

A sequence of actions that is meaningful is more preferable for the construction of jokes than a sequence of unrelated actions.

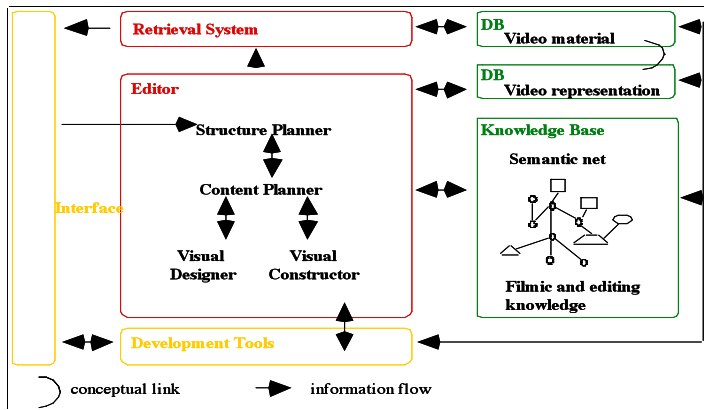
H-Strategy 24

A relationship between two oppositional characters should be established in such a way that the goal of one character is to interrupt the goal of the other in such a way, that is unexpected by the second character. The reaction of the second character must then be influenced by the first so that the second character's mood is downgraded or he suffers in some way. (Mischief + Schadenfreude).

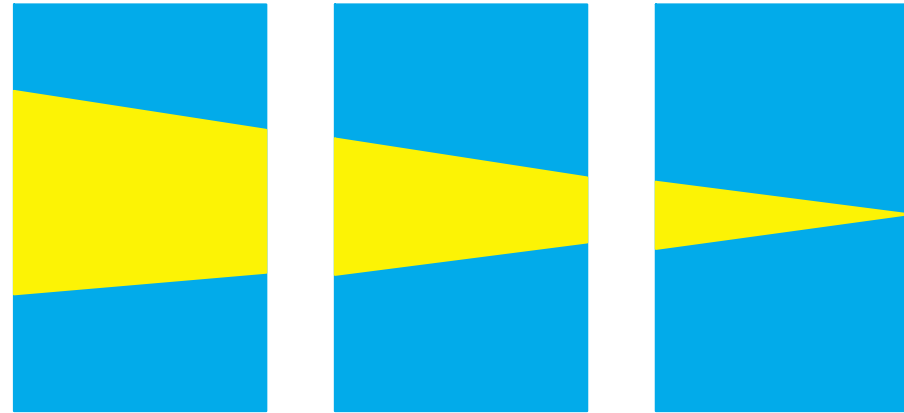
AUTEUR – Content Planner



AUTEUR – Content Planner



Motivation => Realisation => Resolution



■ available video material
 ■ scene relevant material

Motivation	Realisation	Resolution
[approach]	[[search_money, insert_money+], [wait]]	[look_change, take_cup+, leave]
[]	[[process],[provide_cup+]	[]

AUTEUR – Semantic Descriptions - Action

Frame

Name	slip
Domain	motion
Nature of location	outdoors
Set of objects	[banana_peel, dog_shit, soap, ice]
Body part / related object	[shoe]
Location	[road]
Relation Location -> Object	under
Relation Object -> Body part	under
Intention	[unintentional]
Result actions	[sit, lie, kneel, shake, look_back]
Result mood	[anger, rage, astonishment]

AUTEUR – Semantic Description – Event and Concept

Script (to meet)

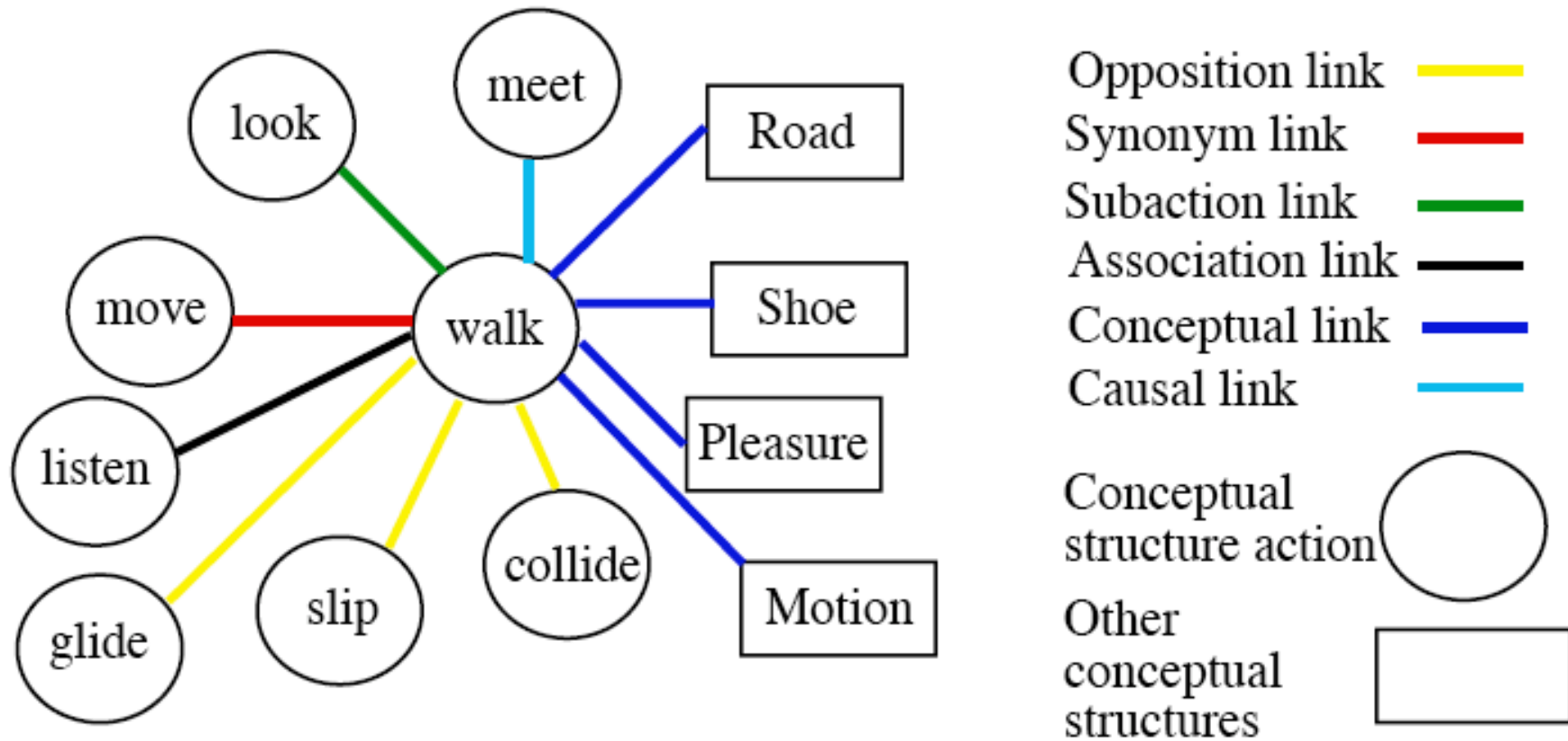
Name	Actor number	Gender	Intention	Motivation	Realisation	Resolution	Episode
meeting	2	any any	meet	[walk] [wait]	[look at] [look at]	[shake_hand] [shake_hand]	date

Conceptual dependency

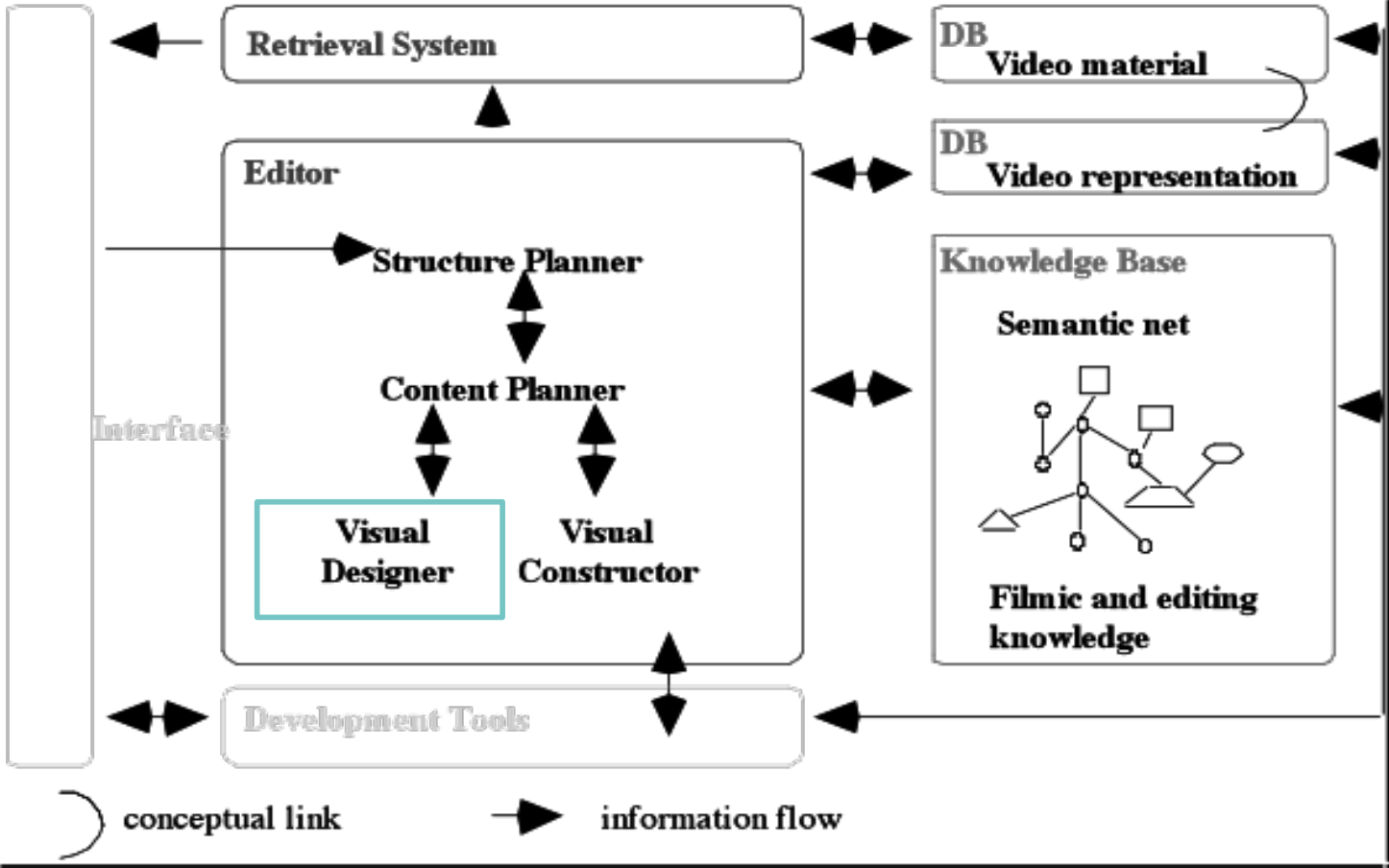
Emotional class name	Body part	Action
pleasure	Head	[lip, up]
	--	whistle

Abstract concept name	Representation structure [character/object, action]
time	[[shadows], [passing]]

AUTEUR – Semantic Description – Semantic Graph



AUTEUR – Visual Designer



AUTEUR – Video Content Representation I

Description Scheme (Character)

Name	Description
Shot ID	
Startframe	
Endframe	
Identifier	Identifier for a character, e.g. a name or a number
Gender	male, female, hermaphrodite, artificial
Age	e.g. young, old, 25, etc.
Race	e.g. black, white, Asian, etc.
Appearance	a structure including: role e.g. lawyer, plumber, stewardess, etc. Costume kind e.g. business suit, apron dress, overall, etc. colour a doublet list providing the major colour for the top and bottom part e.g. [black, white] appeal e.g. casual, formal, etc.

AUTEUR – Video Content Representation II

Description Scheme (Denotative Video Aspects)

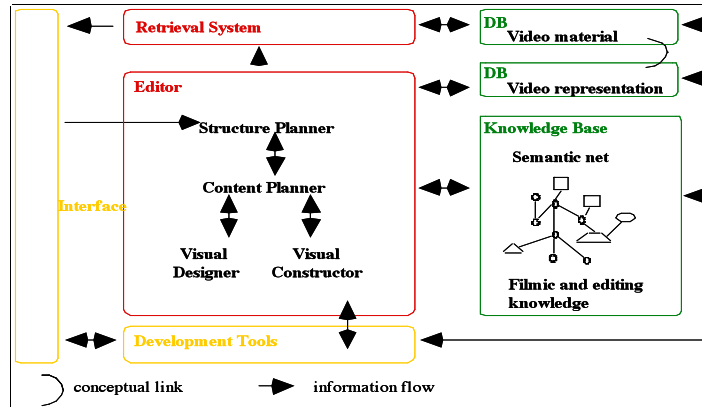
Name	Description
Shot ID	
Startframe	
Endframe	
Identifier	Identifier for a character, e.g. name or number
Relative Position	(Screen position first frame, screen position last frame), e.g. (left, right), (left, middle), (right, right), etc.
Action	e.g. eat, drink, walk, read, etc.
Speed of action	e.g. slow, medium, fast
Direction of action	left, up-left, up, up-right, right, down-right, down, down-left, front, back, circular
Bodygesture	a structure containing: full body horizontal, vertical, left-diagonal, right-diagonal Head profile right, left, half-left, half-right movement up-down, left-right, up, down, left right, circle eyebrows up, down, straight, etc. line of sight left, right, straight, up, down, etc. mouth up, down, straight, open Hand left action/related object e.g.(tap/table) right action/related object e.g.(holding/head) Foot left e.g. tap, lift, etc. right e.g. tap, lift, etc.

AUTEUR – Video Content Representation III

Description Scheme (Media Dependent Aspects)

Name	Description
Shot ID	Identifier
Shotlength	in frames (25 frames for a second)
Startframe	
Endframe	
Shot kind	<p>a structure including:</p> <p>lens movement</p> <p> zoom-in [start camera dist., end camera dist.]</p> <p> zoom-out [start camera dist., end camera dist.]</p> <p> masking left, middle, right</p> <p>lens state (deep focus, foreground-focus, background-focus)</p> <p>camera distance (extreme close-up, close-up, medium, medium long, long, extreme long)</p> <p>camera movement (pan_left, pan_right, tilt_up, tilt_down, roll_left, roll_right)</p> <p>camera position (left, middle, right)</p> <p>camera angle (overhead, high-angle, eye-level, low-angle)</p> <p>film speed (slow motion, normal, fast motion)</p>
Shot colour	<p>colour list of the dominant colours</p> <p>black & white</p>
Shot granularity	fine, medium, strong
Shot contrast	high, medium, low

AUTEUR – Visual Designer



E-Strategy 1

If sequence.kind = Motivation
then

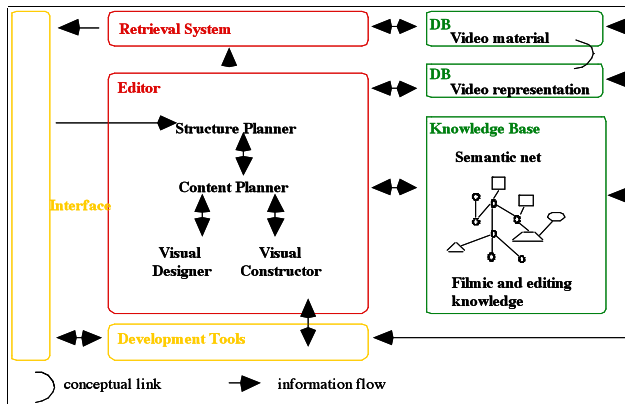
Camera distance of Shot to be chosen is long
=> location.function = outdoor
medium long or medium
=> location.function = indoor

AUTEUR – Video spatial relations between frames



Shot A	Shot B	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) extreme close-up		X	X					
(2) close-up		X	X	X	X		○	
(3) medium close-up			X	X	X			
(4) medium			X	X	X	X	X	○
(5) medium long				X	X	X	X	
(6) long			○			X	X	X
(7) extreme long					○		X	X




AUTEUR – Visual Designer







Value of camera device	Level of content <u>detailness</u>
extreme close-up	object:subparts[form, <u>colour</u>] character: subparts one detail of either head, hand, feet
close-up	object:Instance shape character: Instance or parts of either Head+Id, Hand or Foot
medium close-up medium	object:Instance shape character: Instance Appearance, Head
medium long long	object:Instance shape character: Instance Appearance setting: Time, Location, Lighting
extreme long	object: Instance shape setting: Time, Location, Lighting

AUTEUR – Video spatial relations between frames II

Legend:

Symbols:  shot,  cut,  subject

Type of spatial relations:

-  line,
-  triangle or half circle,
-  square or circle,
-  over cross

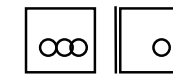
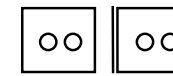
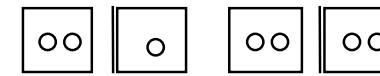
Spatial relationships in the shot and created via a cut



based on the decomposition (1):
hierarchical knowledge
representation of subjects, the
decomposition starts
on parts level :

Class
|
Subclass
|
Instance
|
Parts
|
Subparts

decomposition (2 - n):
the content of each shot
should present the same
hierarchical level within
the knowledge
representation, e.g
Instance - Instance, or
Parts - Parts.



break up into
sub-groups
of the kind
described
to the left and
then use
their
configuration

1 subject

2 subjects

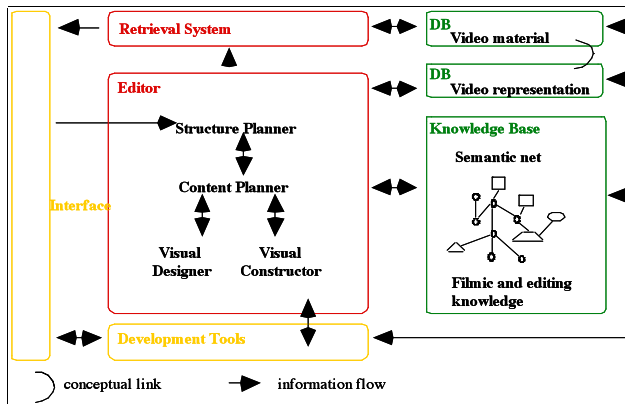
3 subjects

4 subjects

n subject
(where n > 4)

number and size of
characters to be
portrayed and their
spatial relationships
involved.

AUTEUR – Visual Designer



E-Strategy 6

If a sequence is to be established

*where location of shot A \neq location of shot B
or the sequence is the first sequence to be established*

then

create a memory structure of the spatial relations between all characters of Shot B

Location-Memory-Structure

Start Shot-id
 End Shot-id
 List of structures
 List_of_content_relations
 List_of_used_shots

E-Strategy 9

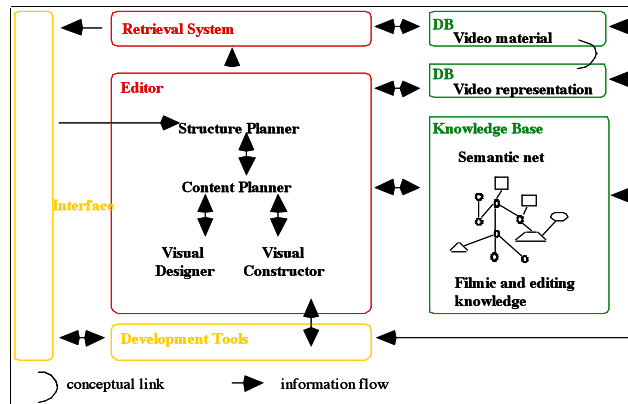
If number of character = 3 and

camera distance of both shots \geq medium long

then

shot A ([left | right]) with shot B ([middle]) => circle / triangle
 shot A ([left | middle]) with shot B ([right]) => circle / triangle
 shot A ([middle | right]) with shot B ([left]) => circle / triangle
 shot A ([left]) with shot B ([middle]) with shot C ([left]) =>
 line or its permutations.

AUTEUR – Visual Designer - Action



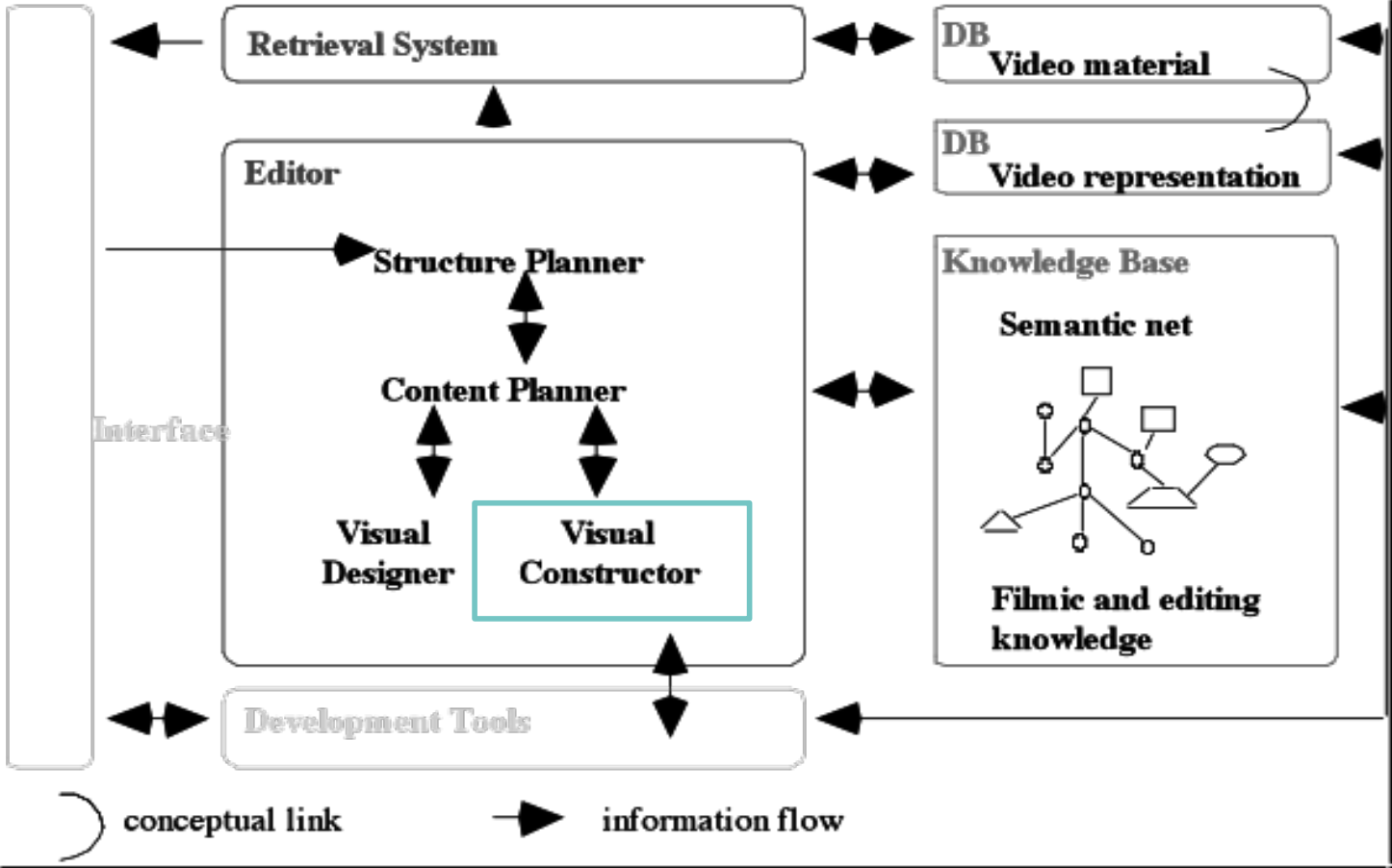
E-Strategy 12

If an action for a character is required and there is no shot available to portray that action then

*isolate the character in a shot
 retrieve the body part related to the action
 retrieve a suitable shot where a body part performs the required action
 build a bridge into or out of this sequence if necessary*


continue with the sequence which was interrupted by this subsequence.

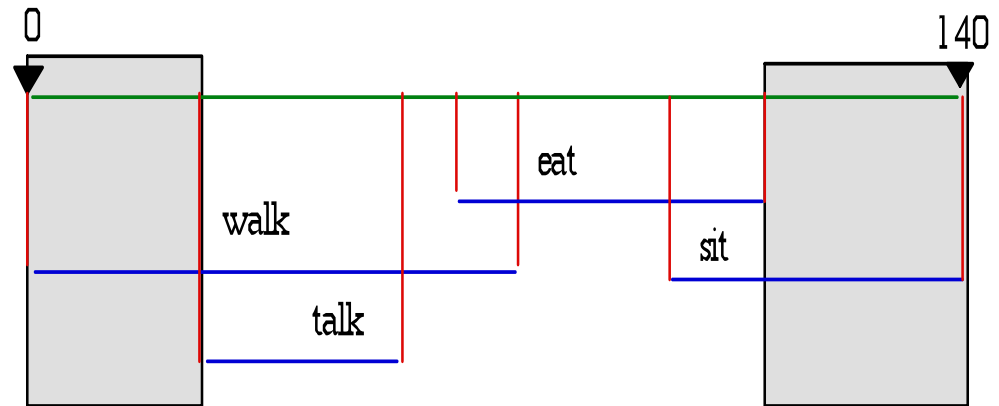
AUTEUR – Visual Constructor



AUTEUR – Video shaping

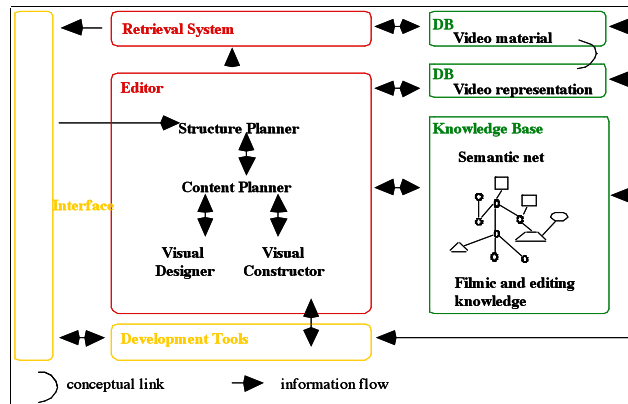



Stratamodel



— Shot
— Annotations

AUTEUR – Visual Constructor



E-Strategy 29

If sequence.action.tempform = contraction and action is a single action

then

favour decomposed forms of presentation where the camera distance of shot A \geq camera distance shot B

E-Strategy 31

If camera distance of a shot \leq close-up and

then

clip it to a length \leq 60 Frames.

E-Strategy 32

If close-up < camera distance of a shot < long and sequence.kind = motivation

then

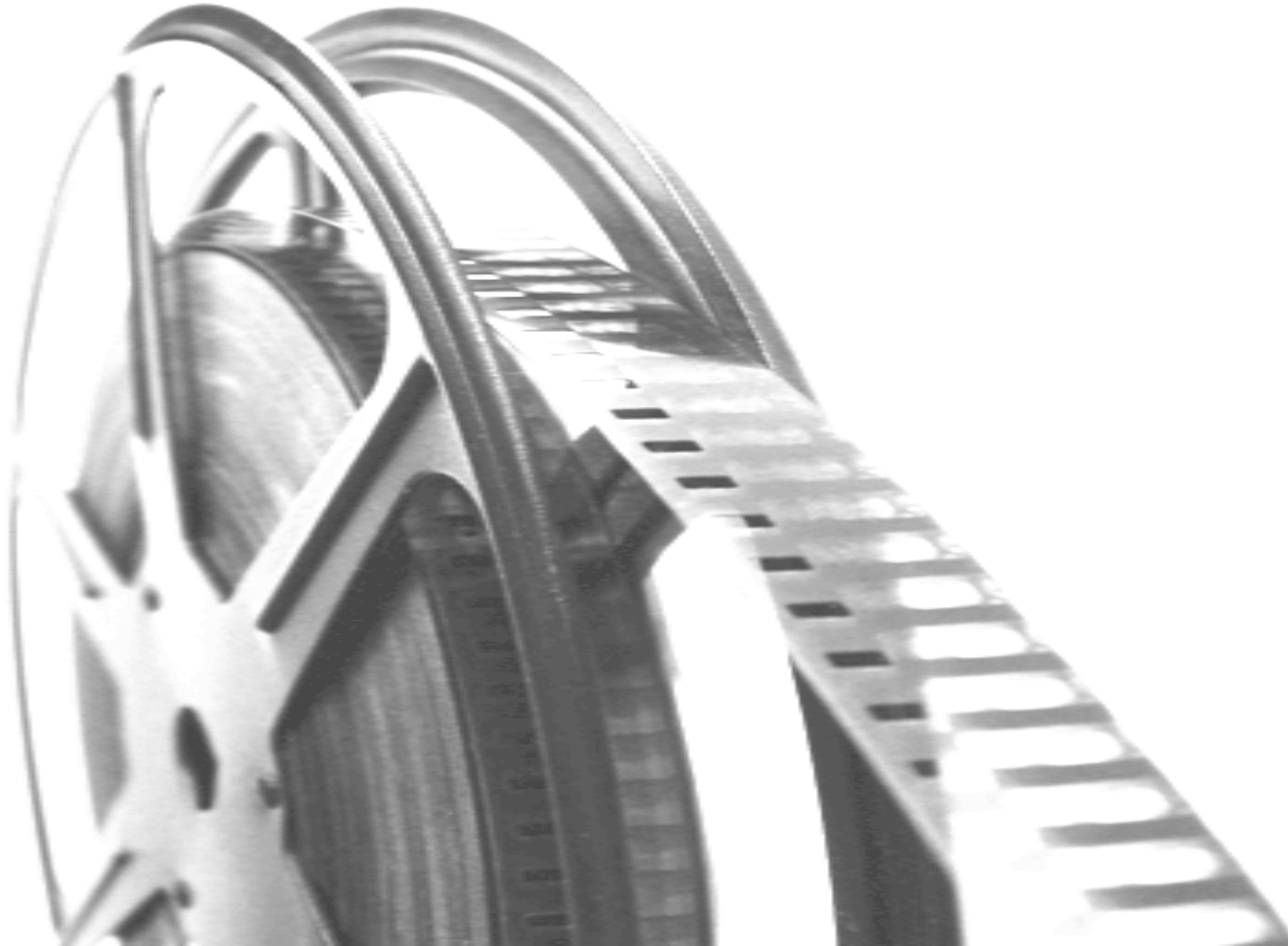
clip it to a length \leq 108 Frames.

AUTEUR – Summary



- Planner approach (backward chaining – goal driven)
- Addresses continuity problem BUT incompletely
- Closed system
- Requires lots of knowledge
- Editing strategies are explicit BUT not final
- Story generation is flexible BUT presentation should be generated rather than using existing material

Video – Additional Applications



Video – Additional Application

Welcome to Terminal Time!

Back Forward Reload Stop Home del.icio.us tag this http://www.terminaltime.com/ Terminal Time

Welcome to the HCS ... UvA - Medewerkers ... Blackboard Learning ... Manuscript Central Google Documenten LEO Deutsch-Englisc... ACM Digital Library caisME Van Dale Taalweb | ... SWI-Prolog's Home

TERMINAL TIME Introduction Description Support Technical Screenings Products Contacts

Project Informa

History is in your hands! Through an audience response-measuring device (applause-meter) connected to a computer, viewing audiences respond to periodic questions reminiscent of marketing polls. These questions occur every 6 minutes during the story. The loudest applause determines the winning answer.

1. What is the most pressing issue facing the world today?

20 A. Men are becoming too feminine and women too masculine.
231 B. People are forgetting their cultural heritage.
C. Machines are becoming smarter than people.
D. It's getting harder to earn a living and support a family.

C. Machines are becoming smarter than people.

Your answers to these questions allow the computer program to create historical narratives that mirror and even exaggerate your biases and desires. Just clap, watch and enjoy. At long last, Terminal Time gives you the history you deserve!

The Terminal Time engine uses the past 1,000 years of world history as "fuel" for creating these custom-made historical documentaries. Each program generated by the machine can be either projected on a screen or broadcast on television monitors. (Although the video and sound are constructed in the computer, the signal is compatible with standard video technology.) Each program lasts approximately 30 minutes.

Audiences experience the project in a similar manner to a movie. A short breakdown of how the audience experiences the work follows:

```

    graph LR
      Intro[2 min. Intro. Similar to Masterpiece Theater Introduction] --> Q1[Q1]
      Q1 --> S1[6 min. Section 1: 1000 to 1750 AD.]
      S1 --> Q2[Q2]
      Q2 --> S2[6 min. Section 2: roughly 1750-1950]
      S2 --> Q3[Q3]
      Q3 --> S3[6 min. Section 3: roughly 1950-present]
      S3 --> Credits[ending credits]
      Credits --> Discussion[Discussion with Audience]
  
```

1. Audience members find their seat and the "movie" begins with a short title sequence and introduction.
2. A series of three questions are asked -- each is shown on screen and read aloud by the digital voiced narrator. The audience answers the questions as a group by applauding. (The answers to these questions changes the story that will follow.)
3. The history lesson begins with a 6 minute section of the movie covering the years 1000-1750.

Done Tor Disabled

Video – Additional Application II

VOX POPULI demo page

http://homepages.cwi.nl/~media/demo/IWA/

Welcome to the HCS ... UvA - Medewerkers ... Blackboard Learning ... Manuscript Central Google Documenten LEO Deutsch-Englisc... ACM Digital Library caisME Van Dale Taalweb | ... SWI-Prolog's Home >>

Vox Populi automatic editing of video documentaries

Introduction

This page presents information about VOX POPULI, a system for automatic generation of Biased Video Sequences. In few words, VOX POPULI uses rhetorical annotations to generate video sequences. Our annotation schema encodes the verbal information contained in the audio channel, identifying the claims the interviewees make and the argumentation structures they use to make those claims. Based on this schema, we construct a semantic graph which is traversed by rhetoric-based strategies selecting video segments.

On-line SMIL demo

The [demo](#) presents a Web interface to generate a short video documentary, as described in the paper. Please [report problems to us](#).

As an example, a prerendered video sequence is [here](#). The Toulmin analysis of the first part of this video sequence (as described in the above-mentioned [paper](#)) is the following:

Two billions dollar bombs on tents

Claim

contradict

Claim

weaken

Concession

support

Claim

I can not think of a more effective solution

War has never solved anything

Other projects using Vox Populi

[VJ Cultuur](#) is the result of a one-year long research about the culture of VJs in the Netherlands and consists of video interviews with 12 Dutch VJs. From the start of this project it was decided to present the video material in an interactive way using Vox Populi.

VOX POPULI in the media

Radio Interviews

[Interview \(in Dutch\)](#) for the Dutch radio program Noorderlicht over scientific developments, broadcasted on national radio. A local copy of the interview is [here](#).

Interview on the Amsterdam radio [Onda Italiana](#) in Dutch, unfortunately it was not recorded.

Press Interviews

[Interview](#) with Elke van Cassel, [Cursor](#) (weekly information magazine of the University of Eindhoven), nr. 13, 2006.

Press Releases

There has been a press release about Vox Populi which has appeared on the following sites:

- [Fact Sheet \(English\)](#) from [CWI](#), the research center I worked for.
- [AlphaGalileo \(English\)](#), independent resource for European research news.

Done

A mock-up of an edited sequence (Real), generated by selecting the position "War in Afghanistan - For" and the point of view "Propagandist - Create Clash"

Examples of not edited interviews

Tor Disabled

Video – Additional Application III

Agent Stories
(1994-1999)

[Kevin Brooks](#)
Writer / Designer

[Glorianna Davenport](#)
Research Supervisor

d e s c r i p t i o n

It is possible to create non-linear stories which allow many different variations of the tale to be interactively played out. Agent Stories is a thinking tool useful for building and ultimately displaying these computationally-based audio/visual narratives.

The writer builds a web of small story "chunks" which are tied together via predefined relationship links. Once this webbed story world or "storybase" is created, the user is able to request a story as constructed by any one of a number of "story agents."

Story agents are software agents with behavioral traits that describe the type of stories they are prone to construct. Although this research is not tied to a particular delivery system, this technique of story definition and construction suggests a model for interactive TV.

The limitation of the program is that story agents can only be as distinct and "interesting" as the complexity of the storybase allows.

c o n t e n t

[Crossing the Street](#) is an evolving and expanding collection of first-person short stories about the effects of taking journeys, large and small.

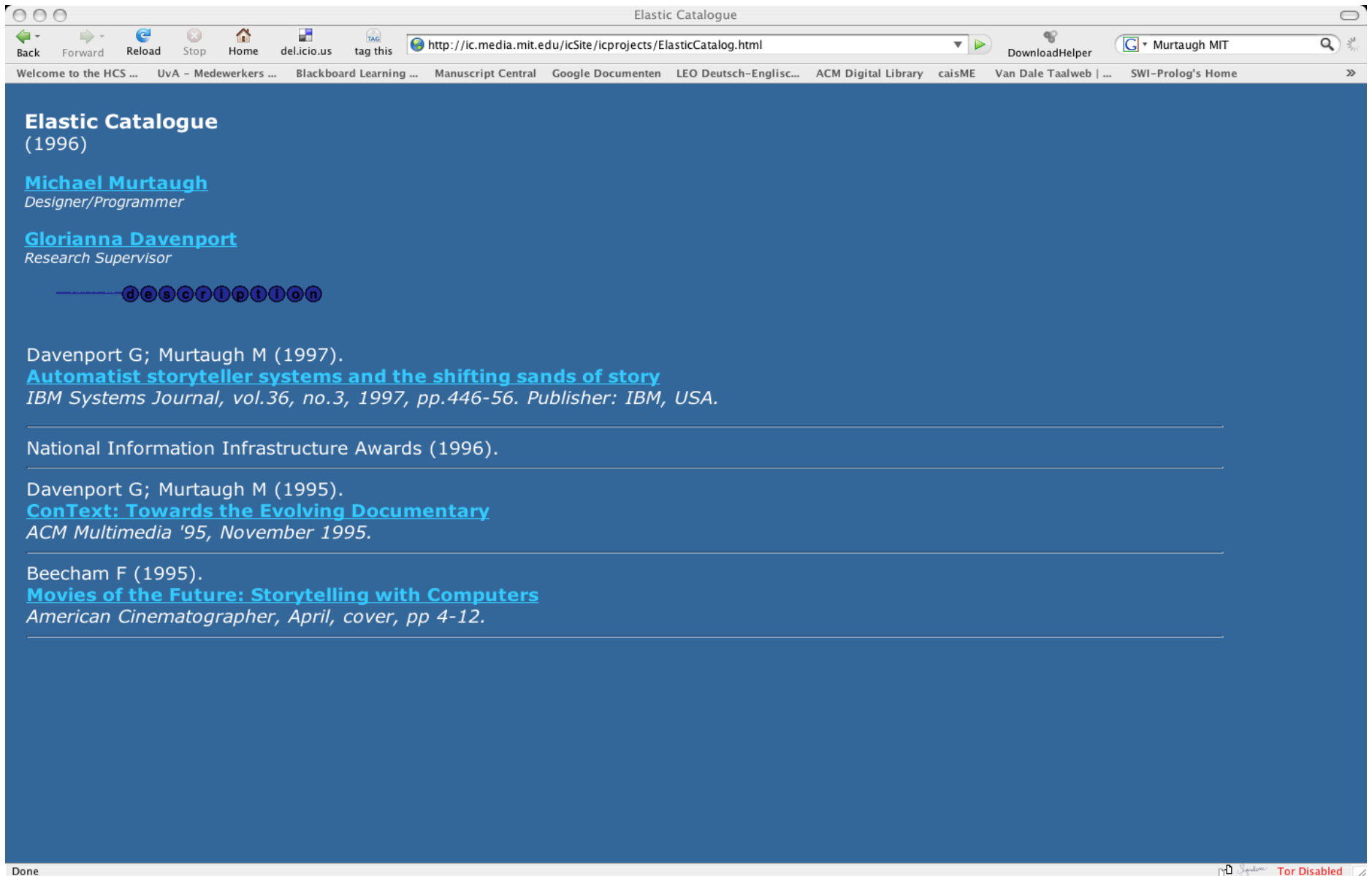
p u b l i c a t i o n s

Davenport G; Bradley B; Agamanolis S; Barry B; Brooks KM (2000).
[Synergistic storyscapes and constructionist cinematic sharing](#)
IBM Systems Journal, vol.39, no.3-4, 2000, pp.456-69.

Brooks KM (1999).
[Metalinear Cinematic Narrative: Theory, Process, and Tool](#)
MIT Ph.D. Thesis.

Brooks KM (1997).
[Programming Narrative](#)
Proceedings. 1997 IEEE Symposium on Visual Languages (Cat. No.97TB100180).

Video – Additional Application VI



Elastic Catalogue
(1996)

[Michael Murtaugh](#)
Designer/Programmer

[Glorianna Davenport](#)
Research Supervisor


— d e s c r i p t i o n —

Davenport G; Murtaugh M (1997).
[Automatist storyteller systems and the shifting sands of story](#)
IBM Systems Journal, vol.36, no.3, 1997, pp.446-56. Publisher: IBM, USA.

National Information Infrastructure Awards (1996).

Davenport G; Murtaugh M (1995).
[ConText: Towards the Evolving Documentary](#)
ACM Multimedia '95, November 1995.

Beecham F (1995).
[Movies of the Future: Storytelling with Computers](#)
American Cinematographer, April, cover, pp 4-12.

Done 

Summary



- **The content determines the application**
- **Content description**
 - **Application dependent**
 - **Complex**
 - **Recourse demanding**
 - **Time critical**
 - **Incomplete**
- **Modular Schemata**
- **Description environment**
 - **Production supportive**
 - **Archive supportive**