SAMT 2006

Readings for the Tutorial: “Human Language Technology for the semantic Annotation of Multimedia Material "

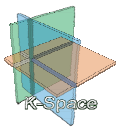
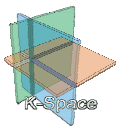


Table of Content

Integration of Modalities/Media

Thierry Declerck (DFKI), with
contributions by Elisabeth André
(University of Augsburg)



Introduction

- Modalities/Media need to be merged at a more abstract level in order to take the maximal advantage of every modality/media involved in a Multimedia (MM) application
- Certain representation formalisms, as they have been defined for NLP, like the well defined technique of unification of typed feature structures, allow to build a semantic representation that is common to all modalities/media involved in a MM application, unifying all the particular semantic contributions on the base of their representation in typed feature structures.

The Role of NLP in Multimedia Applications (E. André)

- Generation of multimedia material including natural language, for the purpose of (coherent) multimedia presentation
- The information contained in the various media has to be very carefully put into relation if one wants to obtain real complementarities of media in the final presentation of the global information.
- Systems for natural language generation have a long tradition with the task of selecting and organising various contributions for the (plan-based) generation of an utterance => a very valuable model for the coordination of media in the context of the generation of multimedia presentations.

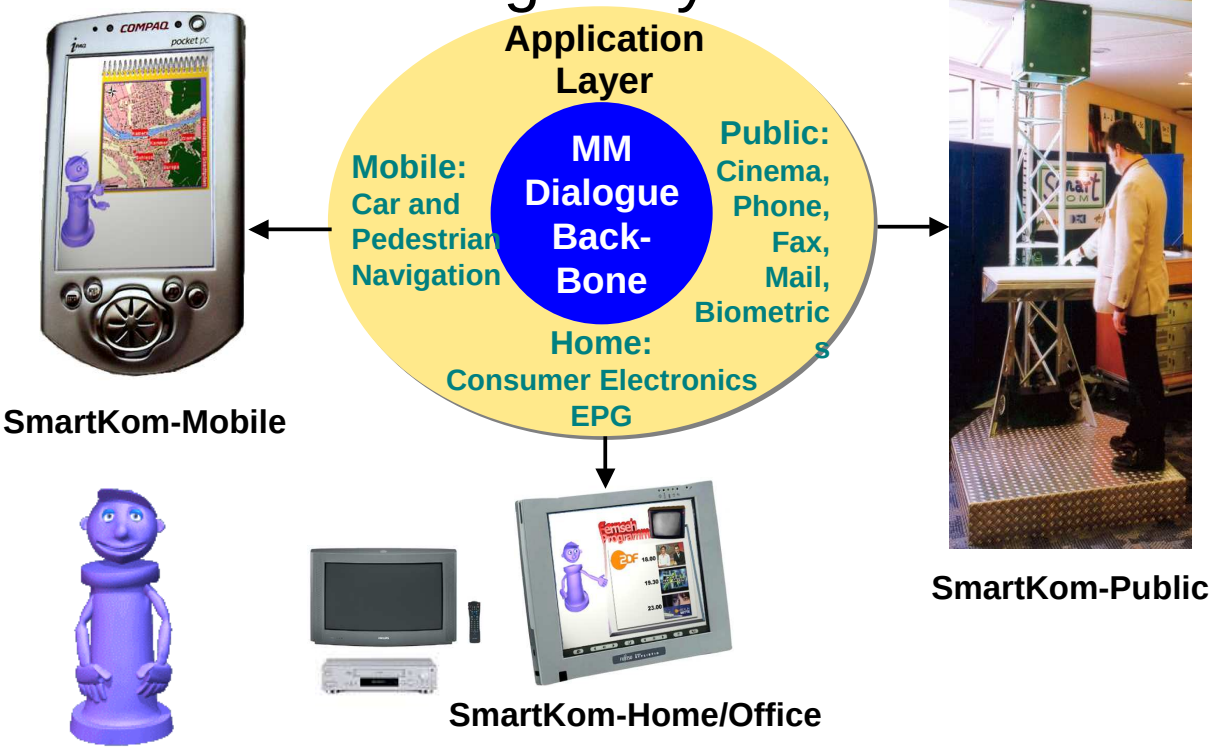
Natural Language Access to Multimedia (E. André)

- Need for fine-grained indexing and retrieval mechanisms allowing users access to specific segments of MM repositories containing specific types of information. NLP can help at various levels:
  - It is first easier to access information contained in the multimedia archive using queries addressed to (transcript of) audio sequences or to the subtitles (if available) associated to the videos as to analyse the pictures themselves (TRECVid)
  - It is further more appealing to access visual data by means of natural language, since the latter supports more flexible and efficient queries as the query based on image features.
  - And ultimately natural language offers a good means for condensing visual information.

The SmartKom MultiModal Scenario:
Semantic Web Technologies for Multimodal User Interfaces.

Prof. Dr. Wolfgang Wahlster, DFKI

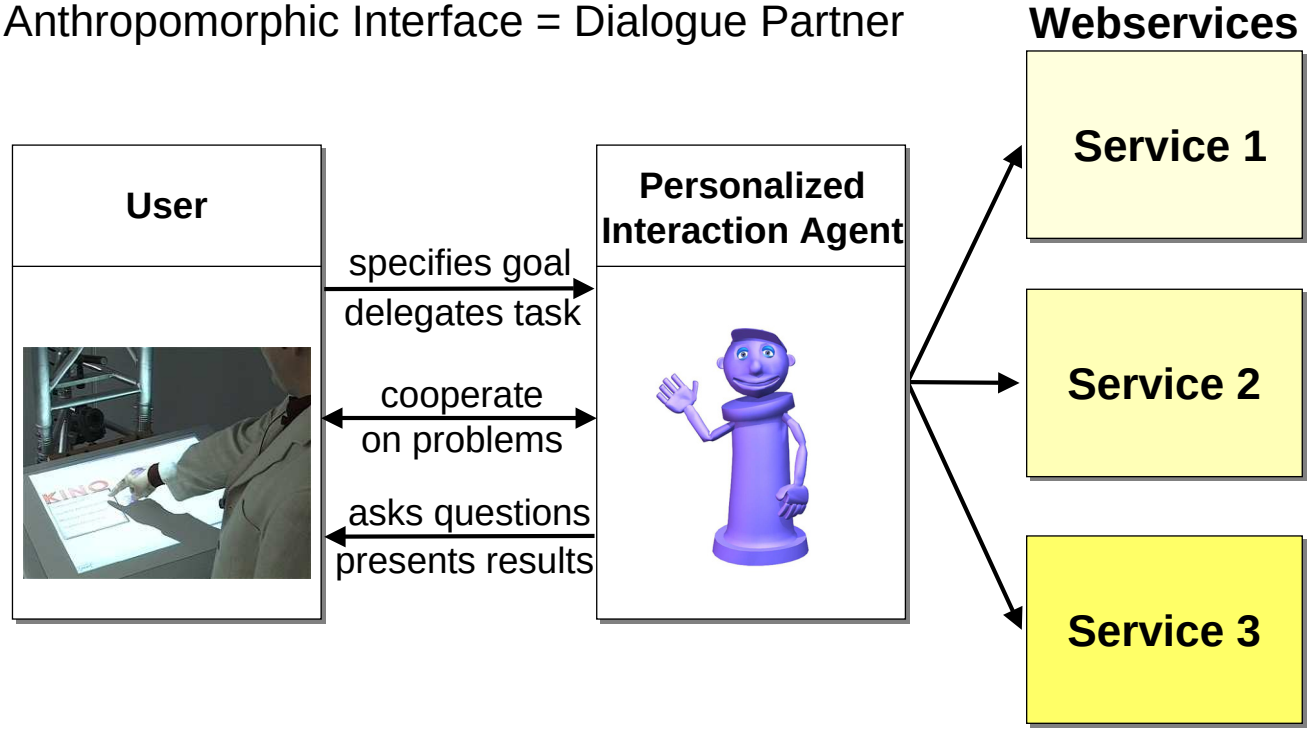
SmartKom: A Highly Portable Multimodal Dialogue System



SmartKom`s SDDP Interaction Metaphor

SDDP = **S**ituated **D**elegation-oriented **D**ialogue **P**aradigm

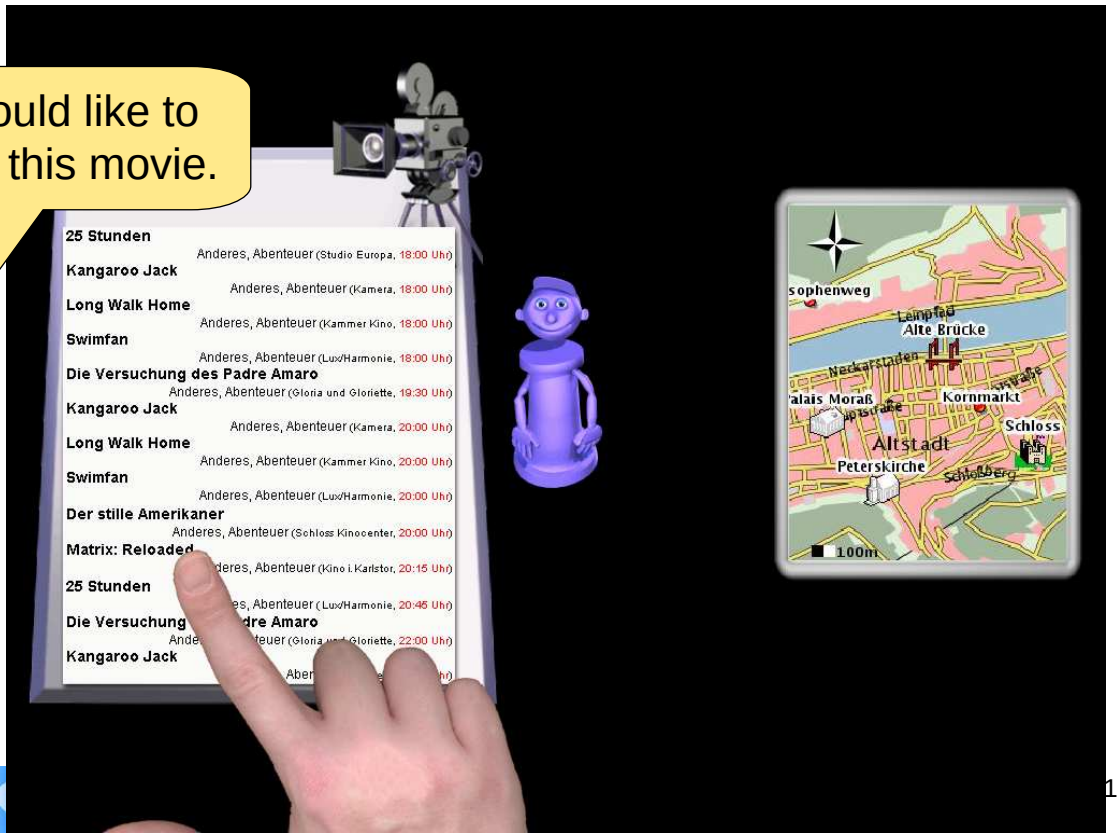
Anthropomorphic Interface = Dialogue Partner



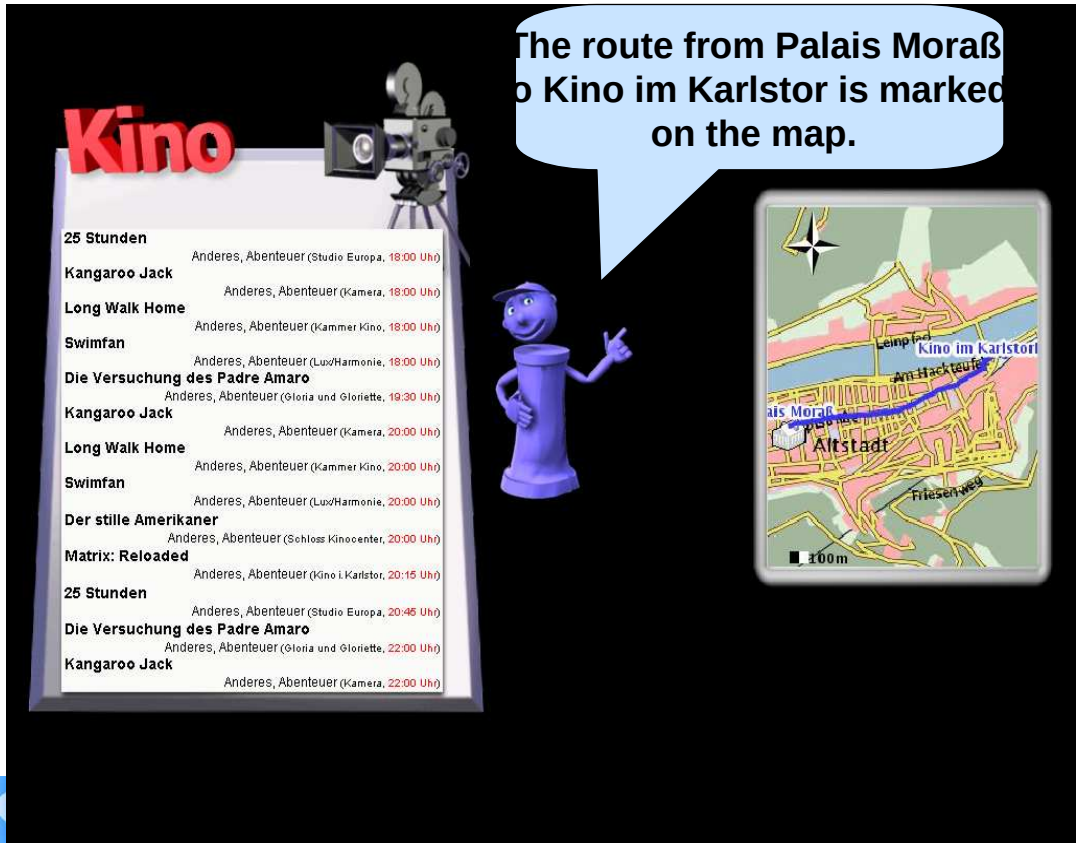
See: Wahlster et al. 2001 , Eurospeech

Generating Maps, Animations and Information Displays on the Fly

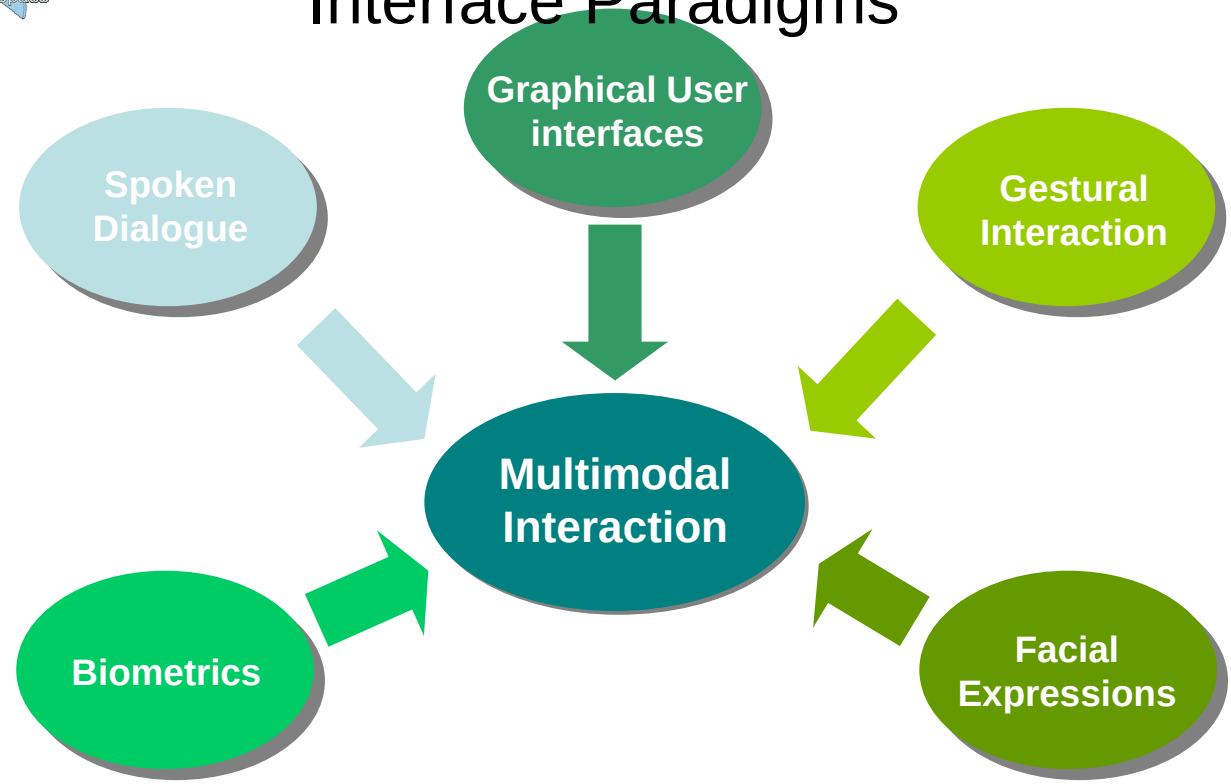
Reference Resolution is based on a Symbolic Representation of the Smart Graphics Output



Synchronization of Map Update and Character Behaviour

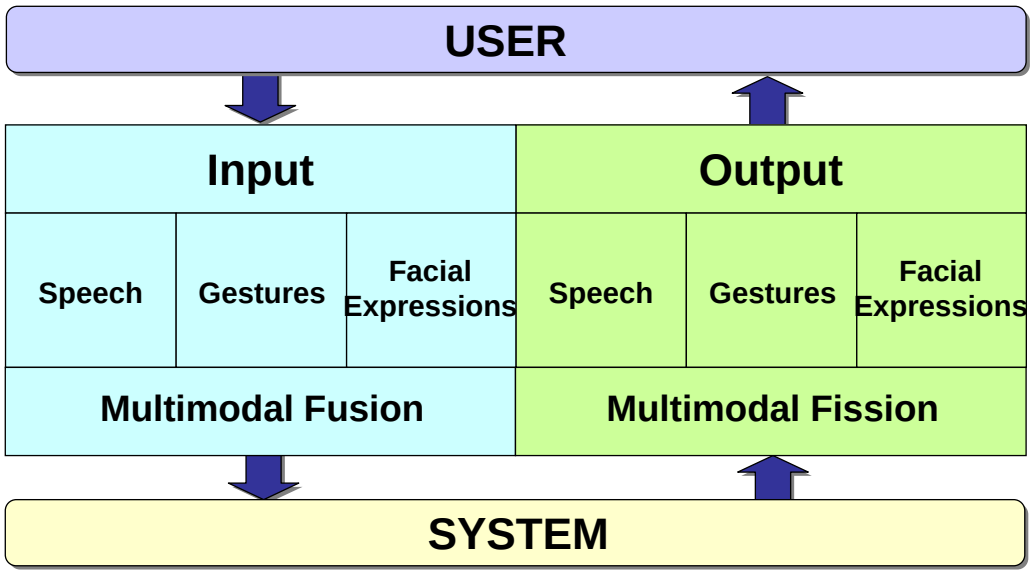


SmartKom: Merging Various User Interface Paradigms



SmartKom: Full Symmetric Multimodality

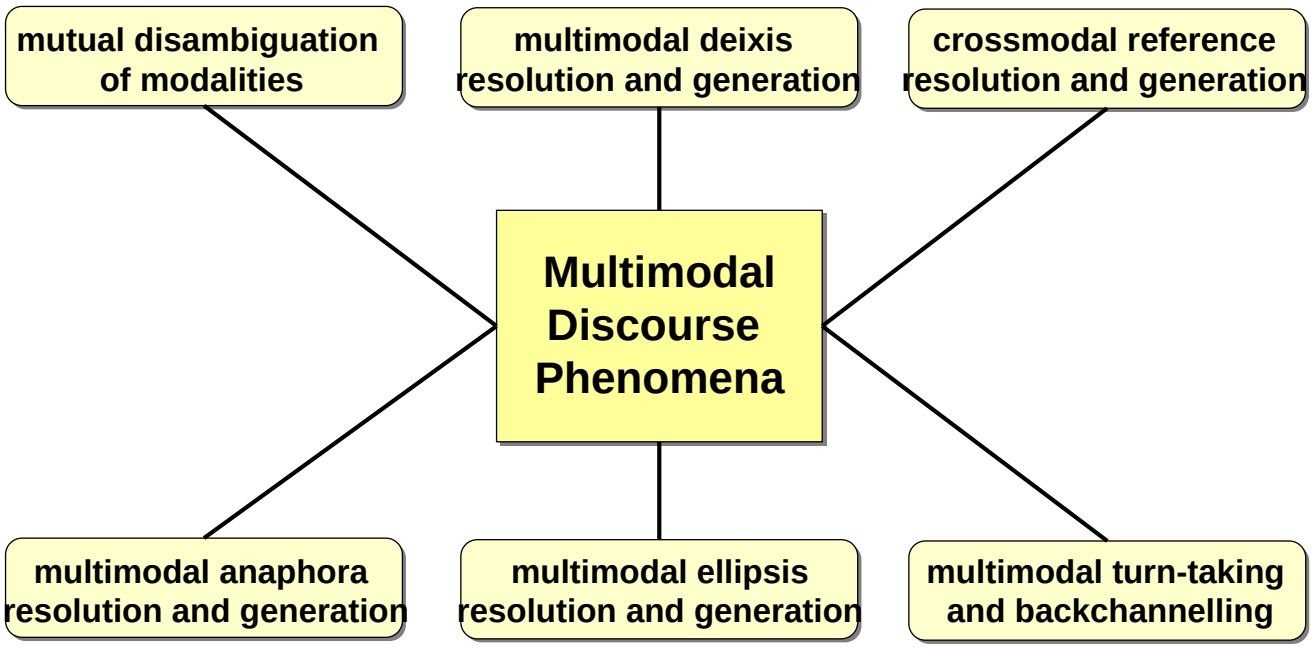
Symmetric multimodality means that all input modes (speech, gesture, facial expression) are also available for output, and vice versa.



The modality fission component provides the inverse functionality of the modality fusion component.

Challenge: A dialogue system with symmetric multimodality must understand and represent the user's multimodal input, AND also its own multimodal output.

SmartKom: Multimodal Discourse Phenomena



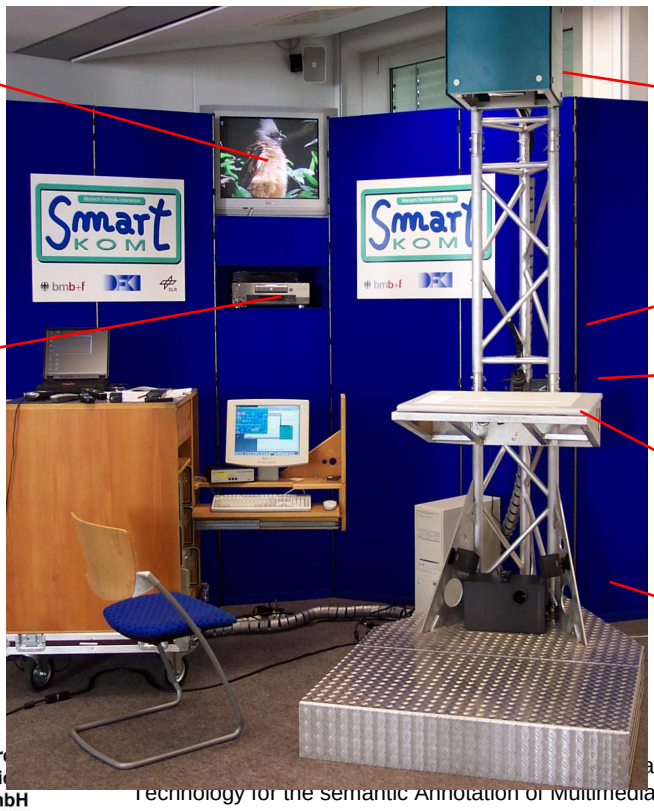
Symmetric multimodality is a prerequisite for processing multimodal discourse .

SmartKom's Multimodal Input and Output Devices

Multimodal Control of TV-Set

Multimodal Control of VCR/DVD Player

3 dual Xeon 2.8 Ghz processors with 1.5 GB main memory



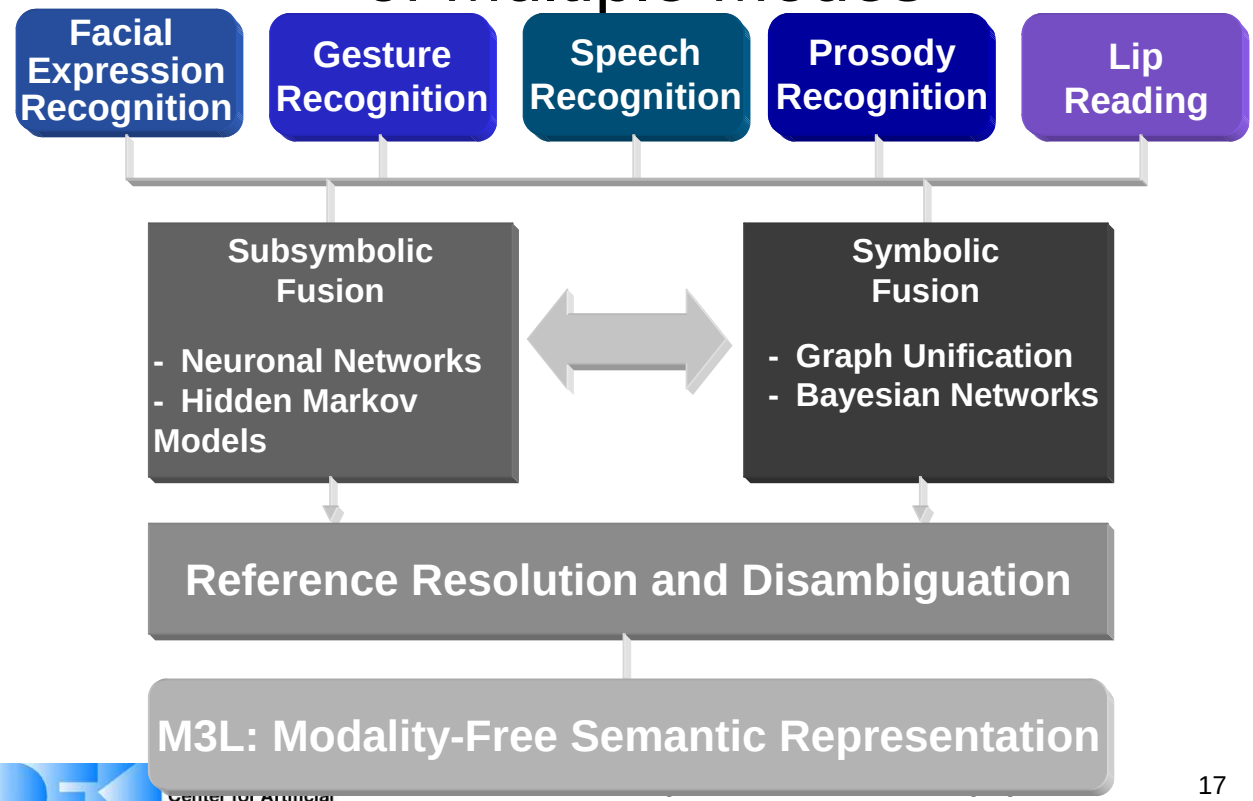
Infrared Camera for Gestural Input, Tilting CCD Camera for Scanning, Video Projector Microphone

Camera for Facial Analysis

Projection Surface

Speakers for Speech Output

Symbolic and Subsymbolic Fusion of Multiple Modes



M3L Representation of a Lattice Fragment

```
<intentionLattice>
[...]
  <hypothesisSequences>
   <hypothesisSequence>
    <score>
     <source>      acoustic     </source>
     <value>       0.96448      </value>
    </score>
    <score>
     <source>      gesture      </source>
     <value>       0.99791      </value>
    </score>
    <score>
     <source>      understanding    </source>
     <value>       0.91667          </value>
    </score>
    <hypothesis>
     <discourseStatus>
      <discourseAction>  set  </discourseAction>
      <discourseTopic><goal>  epg_info  </goal></discourseTopic>
     [...]
     <event id="dim868">
      <informationSearch id="dim869">
       <pieceOfInformation>
        <broadcast id="dim863">
         <avMedium>
          <avMedium id="dim866">
           <avType>  featureFilm  </avType>
           <title>    Enemy of the State     </title>
          [...]
   </hypothesisSequence>
[...]
  </hypothesisSequences>
</intentionLattice>
```

I would like to know more about this

Confidence in the Speech Recognition Result

Confidence in the Gesture Recognition Result

Confidence in the Speech Understanding Result

Planning Act

Object Reference

SmartKom's Computational Mechanisms for Modality Fusion and Fission

Modality Fusion	Modality Fission
Unification Overlay Operations	**Ontological Inferences** Planning Constraint Propagation
M3L: Modality-Free Semantic Representation	

M3L as a Meaning Representation Language for the User's Input

I would like to send an email to Dan.

```
<domainObject>
<sendTelecommunicationProcess>
<sender>.....</sender>
<receiver>.....</receiver>
<document>.....</document>
<email>.....</email>
</sendTelecommunicationProcess
>
</domainObject>
```

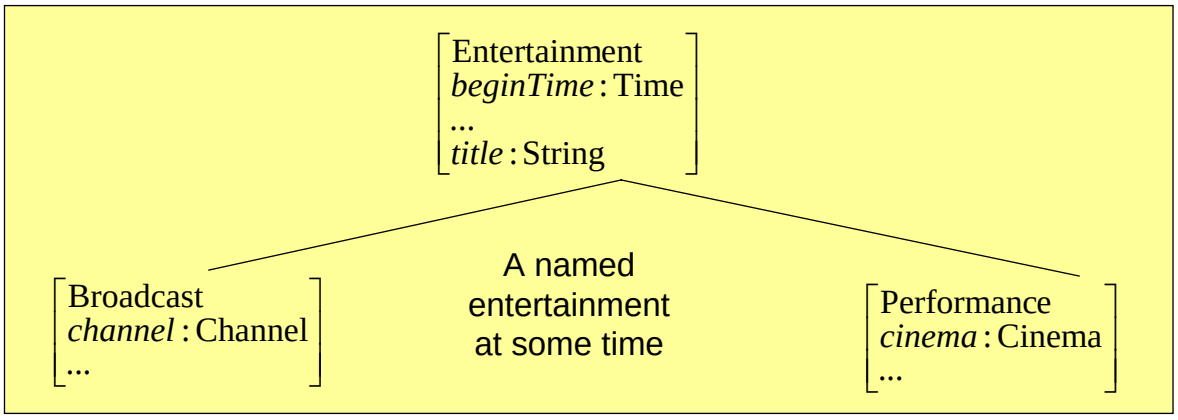
Exploiting Ontological Knowledge to Understand and Answer the User's Queries

Which movies with Schwarzenegger are shown on the Pro7 channel?

```
<domainObject>
 <epg>
   <broadcastDefault>
     <avMedium>
       <actors>
         <actor><name>Schwarzenegger/name></actor>
       </actors>
     </avMedium>
     <channel><name>Pro7</name></channel>
   </broadcastDefault>
 </epg>
</domainObject>
```

```
<beginTime>
 <time>
  <function>
   <at>
    2002-05-10T10:25:46
   </at>
  </function>
</beginTime>
```

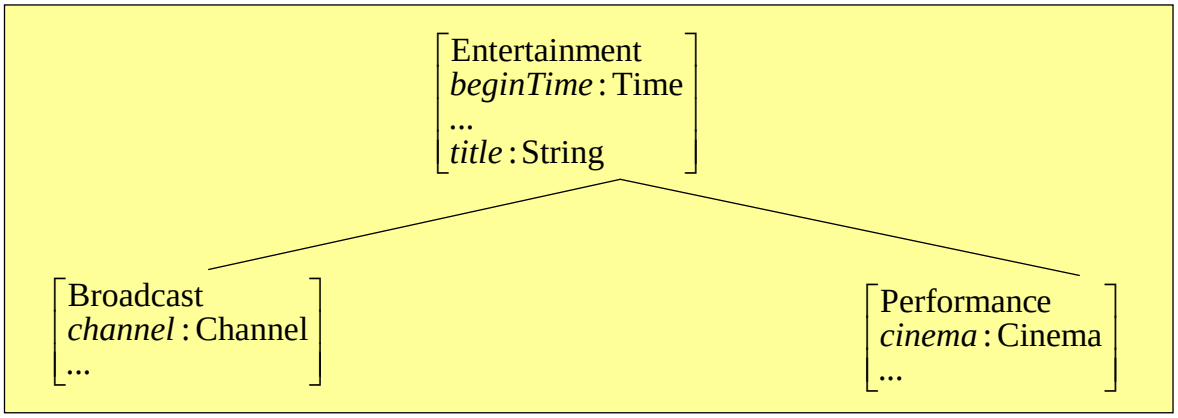

Domain Model: A Type Hierarchy of FS



A named TV program at some time on some channel

A named Movie at some time at some cinema

Unification Simulation



Films on TV tonight

Semantic Web(s) and Language Technology (LT)

Thierry Declerck (DFKI)

Semantic Web(s) and Language Technology (LT)

- Semantic resources as knowledge bases for LT: WordNets (“linguistic ontologies”), Taxonomies (like XBRL for the financial domain), Thesauri (like UMLS for the medical domain) and Ontologies (domain specific or top level).
- All those semantic resources encode their knowledge abstracting over natural language items, terms or expressions (but using natural language for the labelling the abstract items in the knowledge bases - well know problem discussed in philosophy of language).

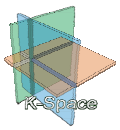
Semantic Web and Language Technology

- Use of Semantic Resources for *Knowledge Markup* of (textual) Web documents -- with authoring tools or supported by automatic text analysis.
- Use of high-level textual analysis for supporting Knowledge Extraction/Learning from (textual) Web Documents.
- A combination of both seems to be appropriate for the implementation of Semantic Web applications
- Some EU projects I am aware of: Esperonto, Sekt ..

Representing Linguistic Information in Ontologies

Paul Buitelaar, Michael Sintek,
Thierry Declerck, Ludger van der Elst, Malte Kiesel

Language Techology Dept. & Knowledge Management Dept. & Competence Center Semantic Web
DFKI GmbH
Saarbrücken, Germany



Overview

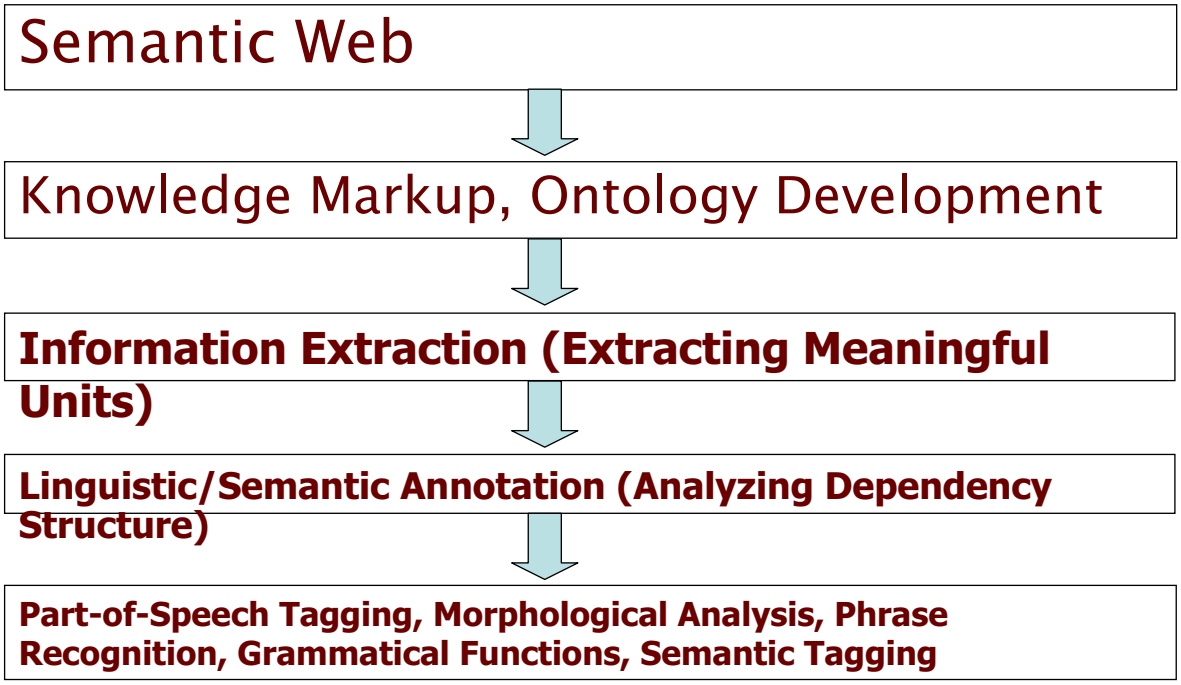
⇒ Ontologies and Linguistic Analysis

Text and the Semantic Web
Ontologies and Linguistic Information
Levels of Linguistic Analysis
Related Work

⇒ Linguistic Information in Ontologies

A Proposal for Modeling LingInfo
Workflow

Text and the Semantic Web



Ontologies and Linguistic Info

Currently, only Use of Labels in RDFS – Not Extensible!

```
<rdfs:Class rdf:ID="Block">
  <rdfs:subClassOf rdf:resource="#Goalkeeper_action" />
  <rdfs:label xml:lang="fr">Bloquer</rdfs:label>
  <rdfs:label xml:lang="en">Block</rdfs:label>
</rdfs:Class>

<rdfs:Class rdf:ID="Cut_down_the_angle">
  <rdfs:subClassOf rdf:resource="#Goalkeeper_action" />
  <rdfs:label xml:lang="fr">Fermer_angle</rdfs:label>
  <rdfs:label xml:lang="en">Cut_down_the_angle</rdfs:label>
</rdfs:Class>
```

Levels of Linguistic Analysis

Lexical Analysis

⇒ Word Category: *Part-of-Speech (incl. Semantic Class)*
⇒ Word Structure: *Morphology*

Phrase Analysis

⇒ Syntactic Units: *Constituency Structure (NP, PP etc.)*
⇒ (Partial) Semantic Units

Dependency Structure Analysis

⇒ Dependency Tree: *Head-Complement/Modifier Structure* (within Phrases or between Phrases)
⇒ Semantic Structure/Phrasal or Sentence Meaning: *Predicate-Argument Structure*

Part-of-Speech, Morphology

Part-of-Speech

⇒ e.g.: noun, verb, adjective, preposition, … (up to 50 tags possible)

Morphology

⇒ Most languages have inflection and declination, e.g.:

Singular/Plural *computer, computers*
Present/Past *reject, rejected*

⇒ Many languages have also complex (de)composition, e.g.:

Flachbildschirm (flat screen) > *flach + Bildschirm*
> *flach + Bild + Schirm*

Phrases

⇒ e.g. nominal - NP, prepositional - PP

NP *a flat screen*
PP *with a flat screen*
NP (recursive) *the computer with a flat screen*
a failure in the motherboard

Dependency Structure

The Dependency Structure (DS) specifies the Role played by a Linguistic Unit (Words, Phrases) within a larger syntactic Context. A Semantic Dependency Structure is the Result of Semantic Analysis which Maps a DS onto Concepts/Relations in a Given Ontology.

The Dell computer that has been rejected was claimed to have suffered from handling.

$$\begin{aligned} &\text{reject}(e_1,x_1,y_1) \ \&\ \text{animate-entity}(x_1) \ \&\ \text{Dell\_computer}(y_1) \\ &\quad \&\ \text{claim}(e_2,x_2,e_3) \ \&\ \text{animate-entity}(x_2) \\ &\quad \&\ \text{suffer\_from}(e_3,y_1,y_2) \ \&\ \text{handling}\ (y_2) \end{aligned}$$

We Propose to Merge/Integrate Dependency Structure with the Ontology

Related Work

⇒ **Relevant Related Work**

SKOS (RDFS Schema for description of thesauri):

http://www.w3.org/2004/02/skos/core/guide/

ISO working group TC37/SC4 on language resource management:

http://tc37sc4.org/

⇒ **Also Related**

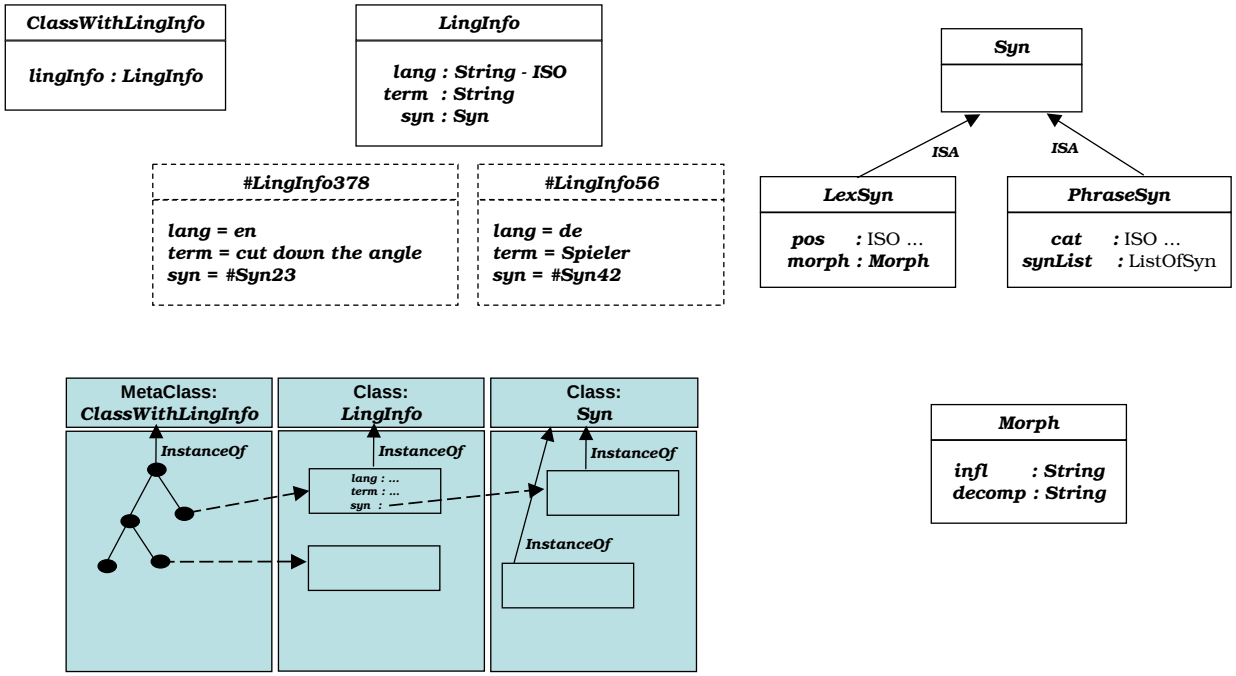
W3C Semantic Web Best Practices WG on WordNet:

http://www.w3.org/2001/sw/BestPractices/WNET/tf

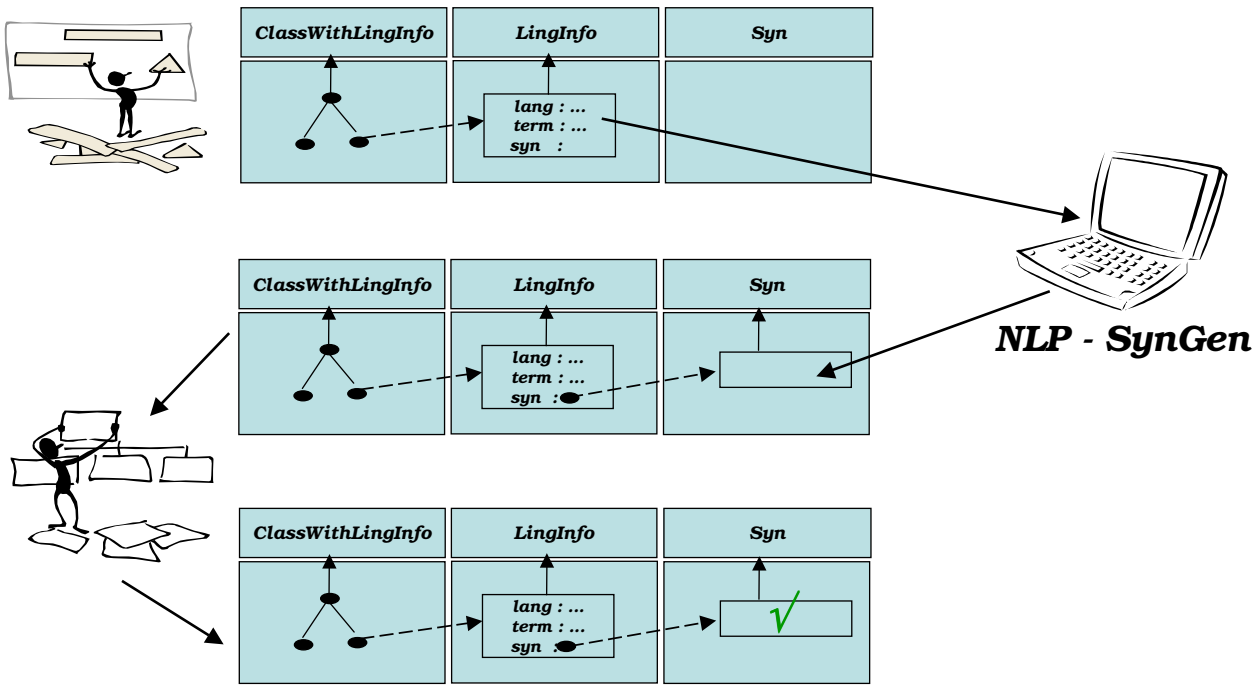
Linguistic Description Scheme in MPEG7:

http://www.i-content.org/mpeg/

A Proposal for Modeling LingInfo

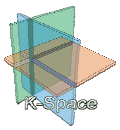


Workflow (Realized in Protege)



Semantics in Multimedia Analysis and Retrieval

Thierry Declerck (DFKI)



State of the Art in video/image processing

- Automatic analysis of video/image material is resulting in so-called “low-level” content features (color, texture, shape etc.).
- Comparing to the way humans perceive and access multimedia content, there is a semantic gap in the field of automatic content detection (and indexing) in video/image processing

Language Technology, Semantic Web and video/image processing

- Need for integration of semantics encoded in associated speech and/or text (superposed text, caption, subtitles, continuous text etc.) or other available modalities (gestures etc.)

Robust NLP Techniques for indexing Multimedia Material

- Automatic Speech Recognition
- OCR processing of text in images
- Keywords analysis in subtitles and captions

=> Generation of single annotation structures for indexing MM material, for example Named Entities

Advanced NLP Technologies for the Content indexing MM Material

- Combination of Language Technology and domain modelling supports the extraction of relevant entities, relations and events from various types of textual documents in various languages.
- Ontology-guided merging of the information extracted from all documents related relevant evens

=> Generation of complex annotation structures, describing relevant (sequences of) events.

Advanced NLP Technologies for the Content indexing MM Material

- Information Extraction: Extraction of entities, relations and events
- Generation of complex annotation structures, describing events (the MUMIS scenario)
- Merging of annotation from different HTML encoded textual sources related to an image: caption, alt, free text (the Esperonto scenario)

MUMIS (July 2000 – December 2002)

- MUMIS was using basic technology for automatic indexing of multimedia material, using data from different media sources (documents, radio and television programmes) to build a specialized set of lexica and an ontology for the selected domain (soccer).
- MUMIS developed and used information extraction techniques and applied them to "co-lateral" texts for extracting significant information (such as the names of players in a team, goals scored...) and use these to build annotations. MUMIS also implemented a merging tool, to combine the event descriptions generated from different data sources, and from sources in different languages (Dutch, English, German).

IE and MM indexing: Experiences in MUMIS

•Technology development to automatically index (with formal annotations) lengthy multimedia recordings (off-line process): Find and annotate relevant entities, relations and events

• Technology development to exploit indexed multimedia archives (on-line process): Search for interesting scenes and play them

Information Extraction

- Information Extraction (IE) is the task of identifying, collecting and normalizing relevant information for a specific application or user.
- The relevant information is typically represented in form of predefined “templates”, which are filled by means of Natural Language (NL) analysis.
- IE combines pattern matching mechanisms, (shallow) NLP and domain knowledge (terminology and ontology).

Information Extraction (2)

IE is generally subdivided in following tasks:

- Named Entity task (NE)
- Template Element task (TE)
- Template Relation task (TR)
- Scenario Template task (ST)
- Co-reference task (CO)

Subtask of IE

- Named Entity task (NE): Mark into the text each string that represents, a person, organization, or location name, or a date or time, or a currency or percentage figure.
- Template Element task (TE): Extract basic information related to organization, person, and artifact entities, drawing evidence from everywhere in the text.

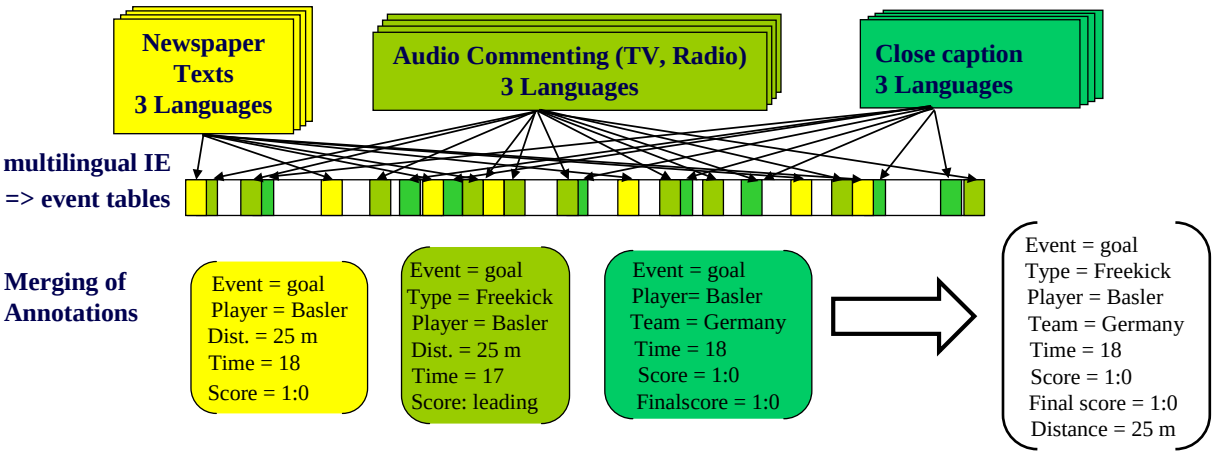
Subtask of IE (2)

- Template Relation task (TR): Extract relational information on employee_of, manufacture_of, location_of relations etc. (TR expresses domain-independent relationships).
- Scenario Template task (ST): Extract pre-specified event information and relate the event information to particular organization, person, or artifact entities (ST identifies domain and task specific entities and relations).
- Co-reference task (CO): Capture information on co-referring expressions, i.e. all mentions of a given entity, including those marked in NE and TE.

Ontology:

```
<lex-element id="ID" concept=„Second-half">
   <... lang="DE" type="main">zweite Halbzeit</term>
   <... lang="EN" type="main">second half</term>
   <... lang=„ES" type="main">reanudacion</term>
</lex-element>
....
```

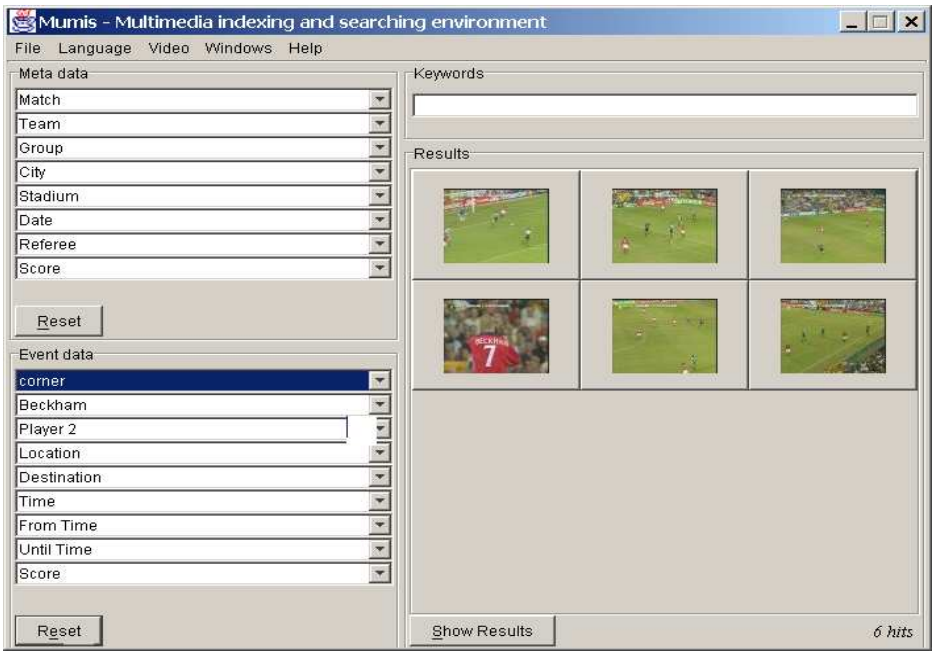

The generated annotation



Events indexed in video recording

•Freekick	•Goal	•Pass	•Defense
•17 min	•18 min	•24 min	•28min
	•1:0		
•Foul	•Freekick		•Dribbling
•Neville	•Basler	•Matthäus	•Campbell
•Basler			•Scholl
•25 m	•25 m	•60 m	

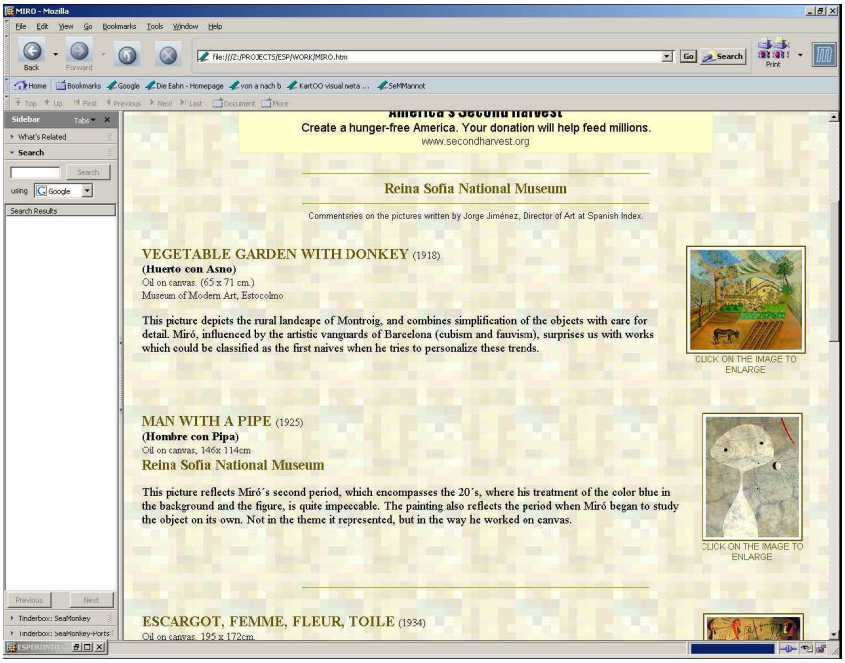
The first user interface of MUMIS



Esperonto (September 2002 – February 2005

- A work package dedicated to the Semantic annotation of still images in Web pages on the base of textual information available around the images. Combination of various html information ("src", "alt", "meta"), with different types of text (caption, title, running text etc.)

Esperonto Scenario



Relevant Text Regions

- Title of the document
- Caption text: „Click on the image to enlarge“ (a non relevant item, to be filtered, also on the base of lexical properties of the words).
- Content of the HTML „Alt“ tag: “'VEGETABLE GARDEN WITH DONKEY'”
- Content of the HTML „Src“ tag: *http://www.spanisharts.com/reinasofia/miro/burro_lt.jpg*
- Abstract text
- Running text

Linguistic Analysis of the Text Regions

- „Alt“ text: 'VEGETABLE GARDEN WITHDONKEY'
  <NP HEAD=“garden” PRE_MOD=“vegetable”
  <POST_MOD CAT= “PP” HEAD=“with”
  NP_COMP_HEAD=“donkey”</POST_MOD>
  </NP>
- Abstract/Running text: “…This picture depicts the rural landcape of Montroig …”
  <SENT SUBJ=“This picture” PRED=“depicts
  OBJ=“the rural lansdscape of Montroig”</SENT>
- Detailled annotation of the direct_object:
  <NP HEAD=“landscape” PRE_MOD=“rural”
  <POST_MOD CAT=“PP” HEAD=“of”
  NP_COMP_HEAD=“Montroig”</POST_MOD>
  </NP>

The Semantic Annotation (1)

- The Toy Artwork Ontology (schematized)
  - Object > Arwtork > Painting [has_creator, has_name, has_subject, has_dimension,has_material, has_genre, has_date...]
  - Person > Artist > Painter [has_name,has_birth_date, part_of_artistic_movement …]

The Semantic Annotation (2)

- The Instantiation
  - Title: Vegetable garden with donkey
  - Creator: Miro
  - Date: 1918
  - Genre: naïve (if correctly extracted by some reasoning on the linguistically and semantically annotated text)
  - Subject: rural landscape of Montroig + garden and donkey (if the association between the title and the explanation given by the art expert can be grouped).
  - Dimension: 65x71
  - Material: Oil on canvas

Direct-Info (January 2004 – December 2005)

- DIRECT-INFO's vision was to create a basic system for semi-automatic extraction of consistent and meaningful semantic information from multimedia content.
- The project developed an integrated system combining the output of basic media analysis modules to semantically meaningful trend analysis results, which shall give executive managers and policy makers a solid basis for their strategic decisions.
- A component was dedicated to the linguistic and semantic analysis of textual documents as XML-encoded dependency structures that comply with the MPEG-7 format for textual description of multimedia content.

Use of the MPEG-7 Standard

- Toward a integrated content indexing framework for annotation resulting from language analysis and multimedia (speech, image, video) processing, using also the whole set of metadata foreseen in MPEG-7 (and MPEG-21), for storing and accessing mutlimedia material.
- Results of language analysis are encoded within the linguistic description schema (LDS) of MPEG-7.
- Also additional to MUMIS is the annotation of positive or negative mentioning of entities mentionned in textual documents, like the name of a soccer team

The MPEG-7 Standard

Thierry Declerck (DFKI), with slides borrowed from Philippe Salembier (UPC, Barcelona)

MPEG-7: Introduction

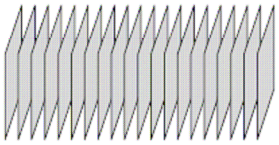
- Allow interoperable search, filter and access of Audio-Visual (AV) content
- Specifies descriptions of features related to AV content as well as information realted to management.
- Define representation of the description wrt syntax and semantics, so that also MPEG-7 descriptions can be generated, for example for consumption

MPEG-7: Some Visual (low-level) Features

- Color
  - Color space, color quantization, dominant colors, scalable color histogram, color structure, color layout, GroupofFrame/Group ofPictures colors
- Texture
  - Homogeneous texture, texture browsing, hedge histogram
- Shape & localization
  - Region-based shape, contour-based shape, region locator, spatio-temporal locator, 3D shape
- Motion & face characterization
  - Camera motion, object motion trajectory, parametric object motion, motion activity

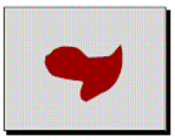
Segmentation in MPEG-7, with corresponding descriptors

Video segments



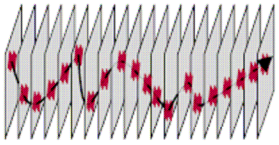
- Color
- Camera motion
- Motion activity
- Mosaic

Still regions



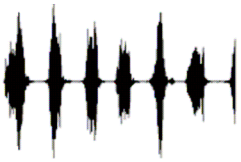
- Color
- Shape
- Position
- Texture

Moving regions



- Color
- Motion trajectory
- Parametric motion
- Spatio-temporal shape

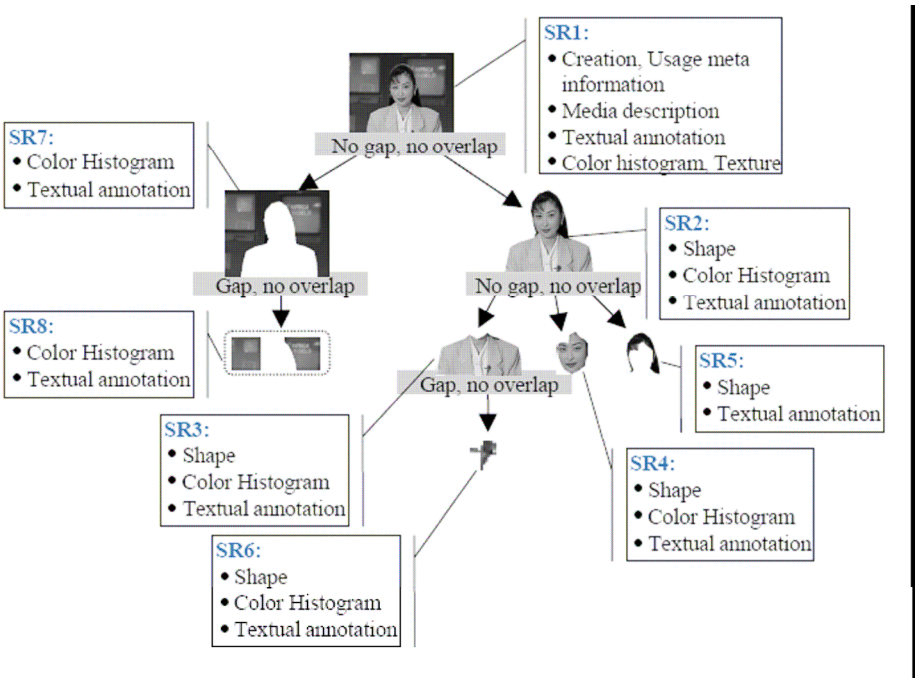
Audio segments



- Spoken content
- Spectral characterization
- Music: timbre, melody

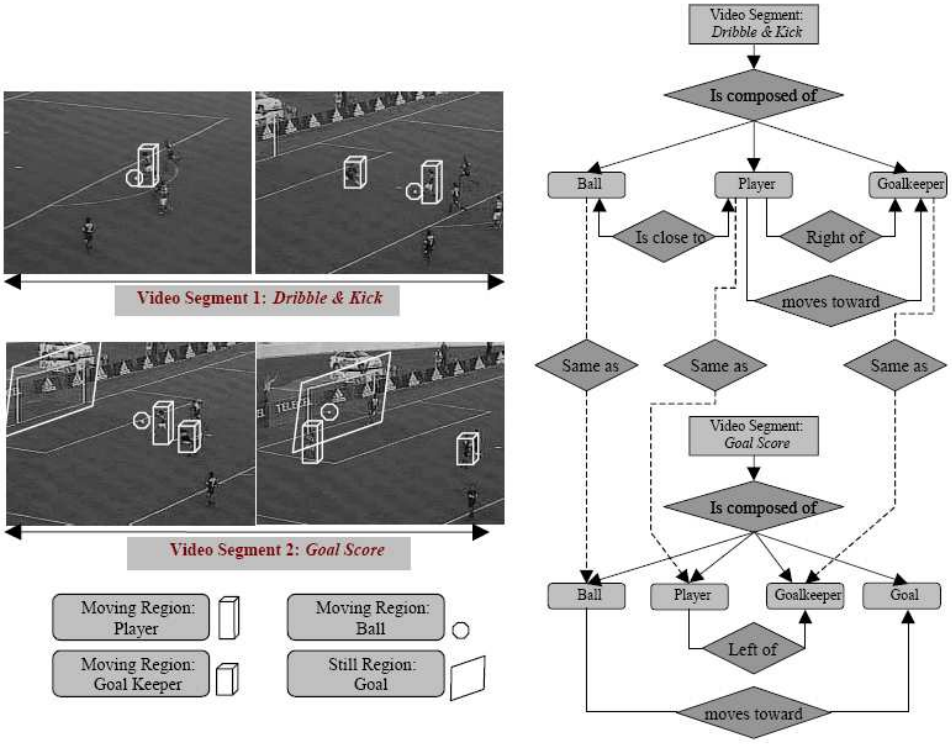
Slide by P. Salembier

MPEG-7: Segment Tree



Slide by P. Salembier

MPEG-7: Graph



The Linguistic Description Scheme (LDS) of MPEG-7

- MPEG-7 foresees 4 kinds of textual annotation that can be attached as metadata to some audio-video content: free text, key words, structured and dependency structure
- The natural language expression used here is "Spain scores a goal against Sweden. The scoring player is Morientes." and the following examples are taken from a former and excellent online tutorial on MPEG-7 by Philippe Salembier.

MPEG-7: Textual Annotation

Free text:

```
<TextAnnotation>
     <FreeTextAnnotation xml:lang="en">
          Spain scores a goal against Sweden.
          The scoring player is Morientes.
     </FreeTextAnnotation>
</TextAnnotation>
```

Structured Annotation:

```
<TextAnnotation>
     <StructuredAnnotation>
          <Who><Name>Spain</Name></Who>
          <WhatAction><Name>score goal</Name></WhatAction>
          <Where><Name>A Coruña, Spain</Name></Where>
          <When><Name>March 25, 1998<Name></When>
     </StructuredAnnotation>
</TextAnnotation>
```

Key-words:

```
<TextAnnotation>
     <KeywordAnnotation>
          <Keyword>score</Keyword>
          <Keyword>Sweden</Keyword>
          <Keyword>Spain</Keyword>
          <Keyword>Morientes</Keyword>
     </KeywordAnnotation>
</TextAnnotation>
```

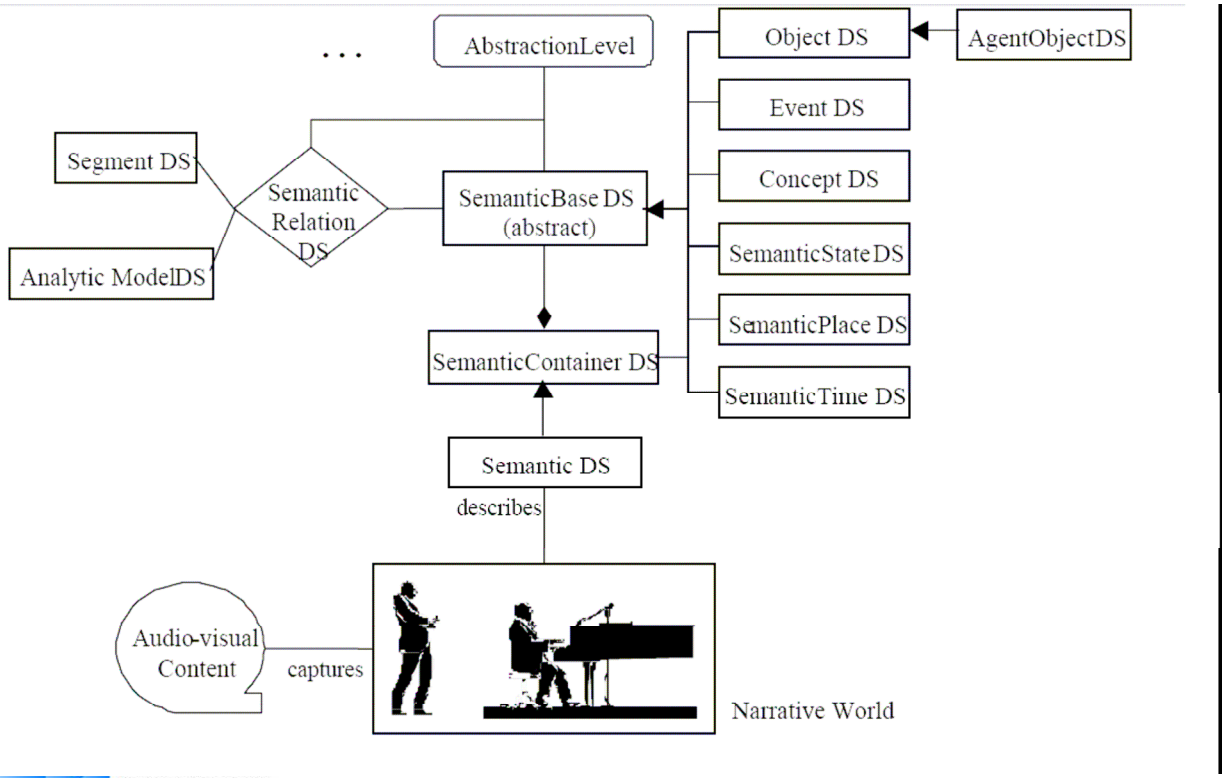
Dependency Structure:

```
<TextAnnotation>
     <DependencyStructure>
          <Sentence>
               <Phrase operator="subject">
                    <Head type="noun">Spain</Head>
               </Phrase>
               <Head type="verb" baseForm="score">scored</Head>
               <Phrase operator="object">
                    <Head type="article noun">a goal</Head>
               </Phrase>
               <Phrase>
                    <Head type="preposition">against</Head>
                    <Phrase><Head>Sweden</Head></Phrase>
               </Phrase>
          </Sentence>
     </DependencyStructure>
</TextAnnotation>
```

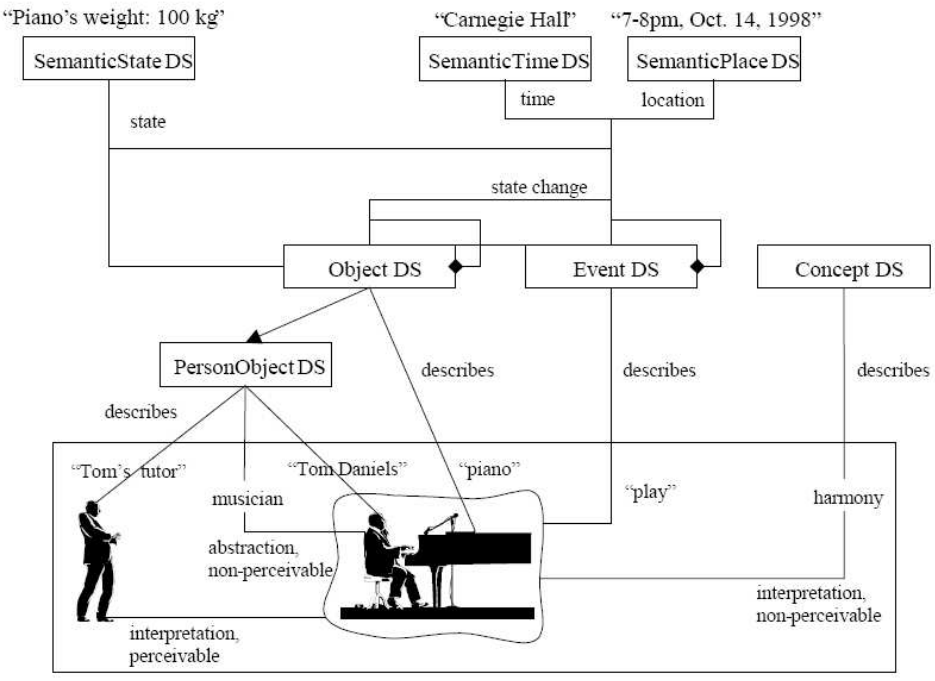

Ex. of MPEG-7 Annotation from Text Analysis

```
<TextAnnotation>
  <WhatObject href="http://www.direct-info.net/mpeg7/cs/LogoCS.2004.xml/di.ta.object.juventus">
    <Name xml:lang="it">bianconeri</Name>
  </WhatObject>
  <WhatAction href=http://www.direct-
            info.net/mpeg7/cs/TextAnalysisCS.2004.xml/di.ta.action.teamMentioned
    <Name xml:lang="it">mentioning of team</Name>
  </WhatAction>
  <Why>
    <Name xml:lang="it">
          Il risultato rappresenta ovviamente un bel viatico per i bianconeri (Juventus) (Juve) , che , a
          dispetto di una stanchezza evidente , possono concedersi alle festività natalizie conil loro
          rassicurante vantaggio di quattro lunghezze , ma non boccia le velleità rossonere perchè
          sono stati proprio i campioni in carica a  fare, come se dice in gergo , la partita , a
          governarla,  a cercare con più insistenza la vittoria.
     </Name>
  </Why>
  <How href="http://www.direct-
            info.net/mpeg7/cs/TextAnalysisCS.2004.xml/di.ta.mentioning.positive">
      <Name xml:lang="it">positive</Name>
  </How>
</TextAnnotation>
```

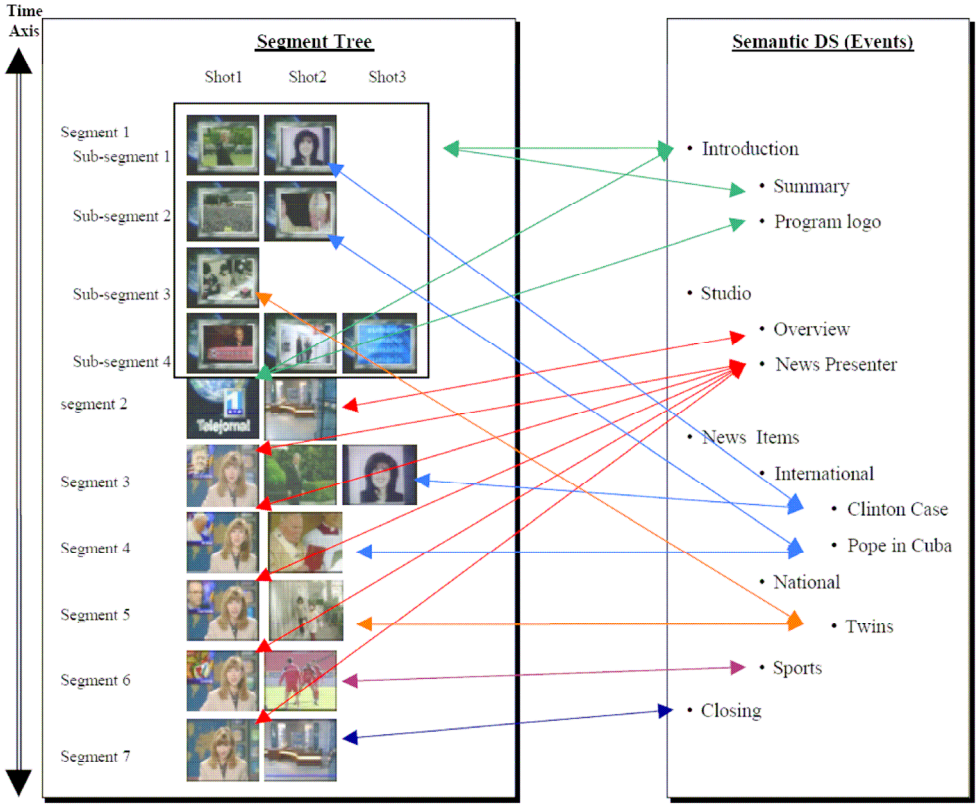
MPEG-7:Semantic Descr. Scheme



MPEG-7:Semantic DS (2)



MPEG-7:Semantic DS (3)



Combining text-based semantic annotation and video stream

- Use other metadata available for video/image content and related metadata detected in textual documents, then
- Use time stamps in the video stream and time information detected in text. Tune by hand (the MUMIS approach).

Relation between NLP and Multimedia

- Before: separated set of annotations (low-level features for MM content and Linguistic-Semantic features for text) put in relations via time codes.
- Future: Merging LL features and LS features via interoperable ontological descriptions of objects and events (Project K-Space, MESH) => Bridging the Semantic Gap in Multimedia Content processing

K-Space

- **Knowledge Space of Shared Technology and Integrative Research to Bridge the Semantic Gap:** K-Space will create a sustainable network of world-leading research teams from academia and industry to conduct integrative research and dissemination activities in semantic inference for automatic and semi-automatic annotation and retrieval ofmultimedia content, aiming at closing the gap between the low-level content descriptions that can be computed automatically bya machine and the richness and subjectivity of semantics in high-level human interpretations of audiovisual media.

Knowledge driven integration/interoperability of key semantic features in both fields

- Work on ontology abstracting over MPEG-7 low-level features already started in aceMedia.
- Availability of a linguistic description scheme (LDS) in MPEG-7 (a subset of it is used for example in the project Busman and Direct-Info). The LDS is supporting the inclusion of textual information available for (segments of) the image/video annotated with MPEG-7.

=> What about adding an ontology description on the top of the LDS, which shares/integrates concepts over the linguistic and the low-level features?

Feature Representation for Cross-Lingual, Cross-Media Semantic Web Applications

Paul Buitelaar, Michael Sintek, Malte Kiesel

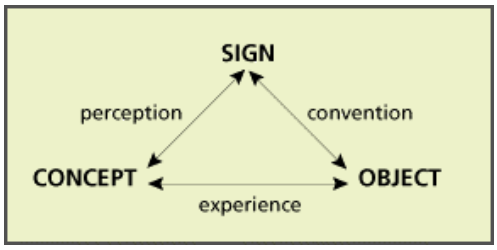
DFKI GmbH
Language Technology Lab, Knowledge Management Department & Competence Center Semantic Web
Saarbrücken/Kaiserslautern, Germany

Overview

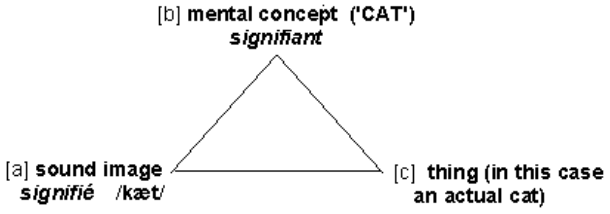
- Motivation
  - Information Extraction from Text & Image Analysis for Knowledge Markup
  - More in General: *Semiotic Triangle* – Multilingual and Multimedia Symbols for Classes & Properties
- Feature Extraction and Representation
  - *Hermeneutic Cycles* - Interacting Layers
  - Possible Application Scenarios
- Ontology-based Feature Representation
  - Proposal (LingInfo / MMInfo)
  - Comparison with SKOS and other Related Work
- Conclusions and Future Work

Motivation – Semiotic Triangle

Ogden & Richards, 1923



- based on Structural Linguistics studies (de Saussure, 1916)



- |. Sowa, 1984)

Motivation – Some Statistics

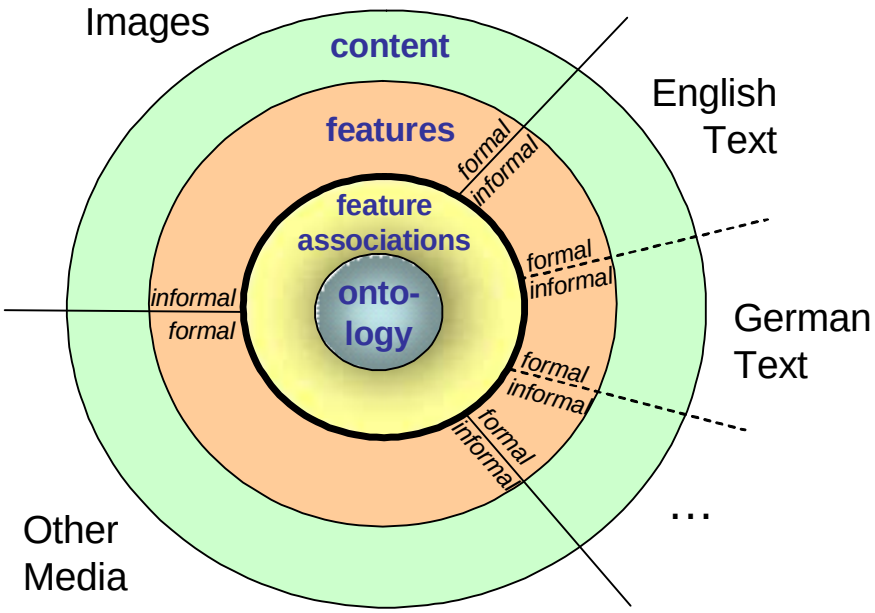
According to OntoSelect (Buitelaar et al., 2004) less than 9% of freely available ontologies represent multilingual terms for classes and/or properties http://views.dfki.de/ontologies

Ontologies in OntoSelect (Nov. 2005)	869
… with Labels with Language Info	72 (8.3%)
EN, EN-GB, EN-US	40
EN FR	13
DE EN	9
EN, ES, FR	3
EN ES	2
EN FI	2
DE	1
NL	1
EL, EN, ES, HI, IA, KO, TR, ZH	1

Features

- Multilingual Features
  - Terms with Linguistic Info and Context Models
  - Example: *goalkeeper*
    - *part-of-speech:* *noun*
    - *morphology:* *goal-keeper*
    - *context (Google hits stats.):* *[gets:420000, holds:212000, shoots:55900, …]*
- Multimedia Features
  - Images with Feature Models
  - Example: *goalkeeper*
    - *color:* *#111111*
    - *shape:* *human*
    - *texture:* *"keypatch-set 223"*

Features – Interacting Layers

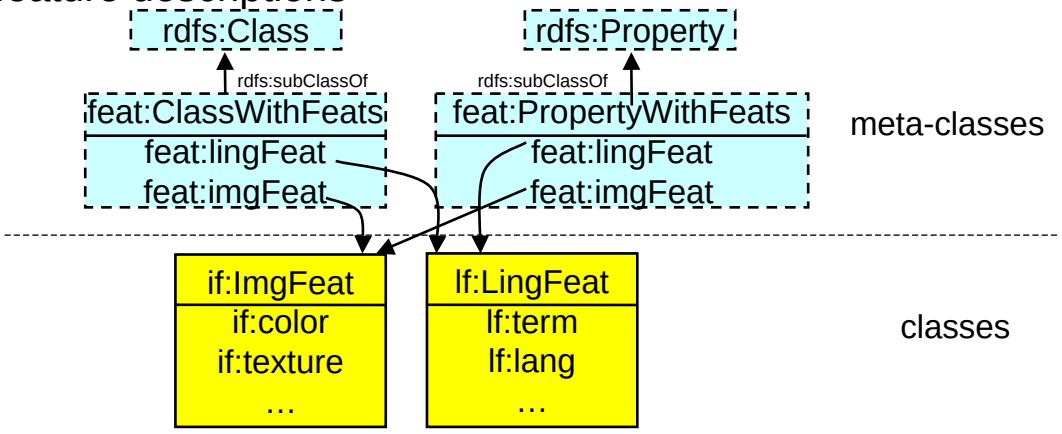


Possible Application Scenarios

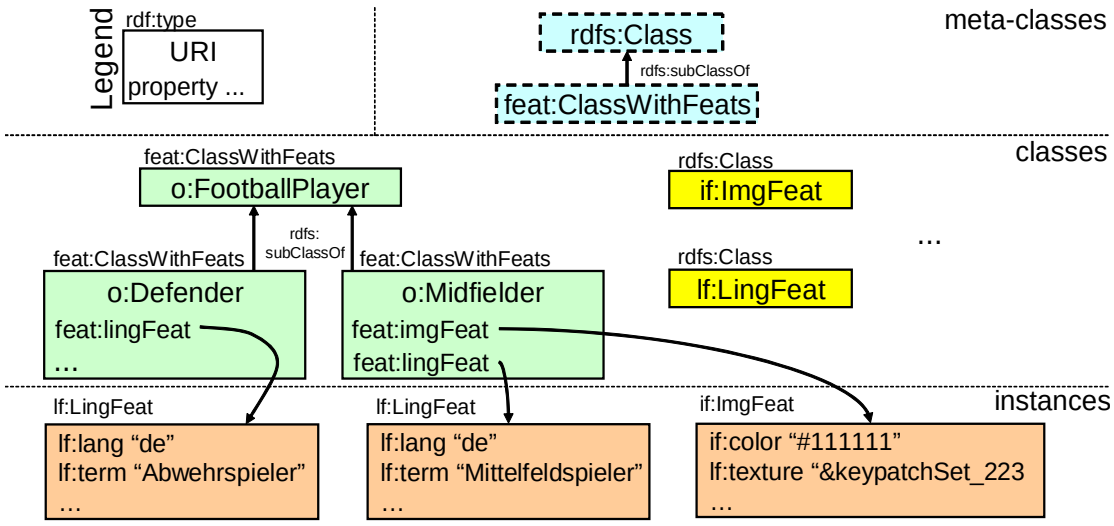
- ***text2image***
  - if we know which terms express a class in English
  - then build a classifier for the classification of images that occur in the context of English terms for this class
- ***image2text***
  - if we know which images represent instances for a specific class
  - then extract German terms for this class from surrounding German text
- ***text2text***
  - if we know which terms express a class in English,
  - and the context features (i.e. words) for these terms
  - and possible translations for these words into German
  - then build a cross-lingual classifier for recognition of unseen German terms for this class
- ***text2class, image2class***
  - if we know which terms express a class in English
  - and the context words for these terms
  - then detect a change in the semantic model for this class by monitoring any change in the context words (similar with image feature models)

Representation – Proposal

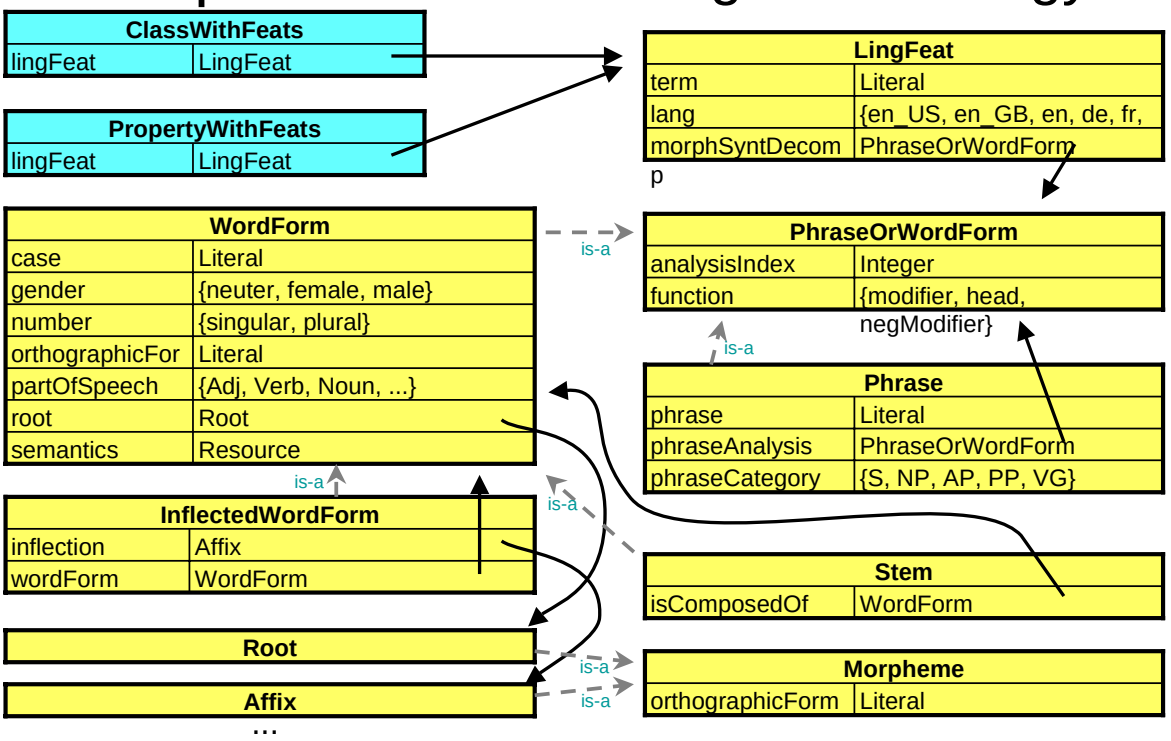
- Attach multilingual and multimedia features to classes and properties (and also instances)
  - use of *meta-classes* ClassWithFeats and PropertyWithFeats with properties lingFeat and imgFeat (with ranges LingFeat and ImgFeat)
- The classes LingFeat and ImgFeat are used for complex feature descriptions



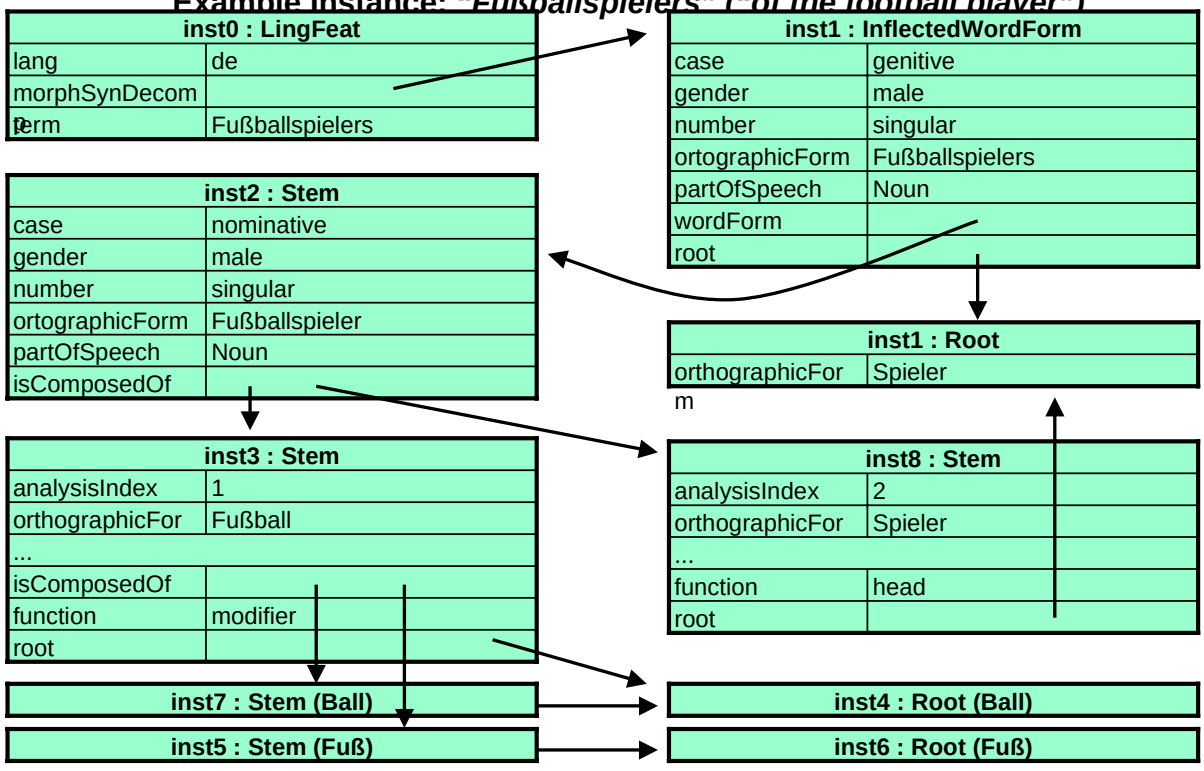
Representation – Simplified Example



Representation – LingInfo Ontology



Example Instance: *"Fußballspielers" ("of the football player")*



Comparison with SKOS

- SKOS (Simple Knowledge Organisation System): RDF-based W3C Working Draft for thesauri, terminologies, glossaries, …
  - Uses sub-properties of **rdfs:label** (**skos:prefLabel**, **skos:altLabel**) plus **xml:lang**
  - Can thus only handle literal-valued linguistic information (= multilingual strings)
  - Uses **foaf:depiction**, **skos:prefSymbol**, etc. for images
    - these do not allow complex image feature representations to be attached to concepts
  - Mixes linguistic and semantic knowledge
    - **skos:broader** and **skos:narrower** have no clear semantics (intentionally)
    - **skos:broaderGeneric** and **skos:narrowerGeneric** with **rdfs:subClassOf** semantics

Related Work

- ISO/TC37/SC4: "Language Resource Management"
  - Builds on earlier EAGLES and ISLE initiatives for standards in linguistic resources (MILE format for lexical entries)
  - Working on the "Lexical Markup Framework" (LMF)
    - XML Format for 'Machine Readable and NLP Lexicons'
- GOLD Ontology for Linguistics
  - Ontology initiative for information organization around endangered languages
- W3C WG "Semantic Web Best Practices" - Task Forces
  - Thesauri Portability – SKOS
  - WordNet – OntoWordNet, RDF/OWL WordNet model
  - Vocabulary Management – Human-Readable annotations for terms