Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

// SMARTWEB

IJCAI Workshop AI4HC, Hyderabad, 6/1/2007

SmartWeb Handheld — Multimodal Interaction with Ontological Knowledge Bases and Semantic Web Services

Daniel Sonntag, Ralf Engel, Gerd Herzog, Alexander Pfalzgraf, Norbert Pfleger, Massimo Romanelli, Norbert Reithinger

German Research Center for Artificial Intelligence
66123 Saarbrücken, Germany

daniel.sonntag@dfki.de

Agenda

- SmartWeb & Multimodal QA Requirements
- Architecture Approach
- Ontology Representation and Web Services
- Semantic Parsing and Discourse Processing
- Conclusions

Project Goals

- SmartWeb goal:

  - Intuitive multimodal access to a rich selection of Web-based information services.

- HCI/Dialogue system goals:

  - Demonstrate the strength of **Semantic Web** technology for information gathering dialogue systems.
  - Show how knowledge retrieval from ontologies and Web Services can be combined with advanced dialogical interaction, e.g., **system** clarifications.
  - Ontology-based **integration** of verbal and non-verbal system input (fusion) and output (reaction/presentation).

Smartweb Requirements

- Multimodal dialogue with question answering functionality.
- Speech is dominant input modality for interaction.
- Multimodal recognition for speech or gestures.
- Modality interpretation and fusion, intention processing.
- Modality fission, result rendering for text, images, videos, graphics, and synthesis of speech.
- Reuse already existing components.
- Control the message flow in the system.

3G smartphone

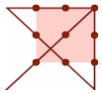
Dashboard display

Motorbike cockpit

The SmartWeb Consortium

Federal Ministry of Education and Research

AIFB



DAIMLERCHRYSLER

European Media Lab

Funded by the German Government and Industry

Funding: 13.7 M €, Budget: 24 M €

Scientific Director: Wolfgang Wahlster

Project Duration: 2004-2008

More than 60 Researchers and Engineers

Fraunhofer FIRST Institut Rechnerarchitektur und Softwaretechnik

Chair for Pattern Recognition FAU Erlangen-Nuremberg

IMS Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

LMU IPSK Ludwig-Maximilians-Universität München

ontoprise SEMANTICS FOR THE WEB

SIEMENS

T-Systems

SYMPALOG Speech Technologies

ISI Berkeley, USA

UNIVERSITÄT DES SAARLANDES

Application Scenarios

- Personal guide at the FIFA Worldcup 2006
  - SmartWeb: *Getting Answers on the Go* (keynote Wahlster, ECAI 2006)

    http://www2.dfki.de/~wahlster/ECAI2006/
  - German Telekom Mobility and Navigation Scenario

    http://smartweb.dfki.de/SmartWeb_FlashDemo_eng_v09.exe

Interaction Guidelines

- Multimodal Guidelines and Assets
  - *Multimodality*: More modalities allow for more natural communication.
  - *Encapsulation:* Encapsulate user interface proper from the rest of the application.
  - *Standards:* Re-use own and others resources.
  - *Representation*: A common ontological knowledge base eases data flow, avoids transformations, and provide a basis for processing natural language dialogue phenomena.

Multimodal Input and Output

Speech/Spoken dialogue over Bluetooth Headset or PDA micro

Gestures on Graphical user interface based on Pen Input

Handheld scenario

Multimodal Interaction

Face Camera/Facial EXpression

Bio Signals

Haptic Feedback/ Physical Action by handlebars on BMW bike

Interaction Example

- U (Query): Show me the mascot of the football WCS.
- S (Clarification): Which year? 2006 | 2002 | 1998 | 1994 | 1990
- U (Feedback): 2006
- S (Multimodal): GOLEO

- U (Query): I need some texts about football rules.
- S (Intermediate Result): Paragraph: Yellow card | Paragraph: Red card | Paragraph: Penalty shot
- U (Feedback): What does red and yellow card mean?
- S (Final Result) Paragraph: Yellow, yellow-red, and red cards are shown …

Handheld Interaction Example

(1) **U:** "When was Germany world champion?"

(2) **S:** "In the following 4 years: 1954 (in Switzerland), 1974 (in Germany), 1990 (in Italy), 2003 (in USA)"

(3) **U:** "And Brazil?"

(4) **S:** "In the following 5 years: 1958 (in Sweden), 1962 (in Chile), 1970 (in Mexico), 1994 (in USA), 2002 (in Japan)" + [*team picture, MPEG-7 annotated*]

(5) **U:** Pointing gesture on player *Aldair* + "How many goals did this player score?"

(6) **S:** "Aldair scored none in the championship 2002."

(7) **U:** "What can I do in my spare time on Saturday?"

(8) **S:** "Where?"

(9) **U:** "In Berlin."

(10) **S:** *The cinema program, festivals, and concerts in Berlin are listed.*

Inducting/Deducing Enumeration question

Ellipsis resolution/ Query completion

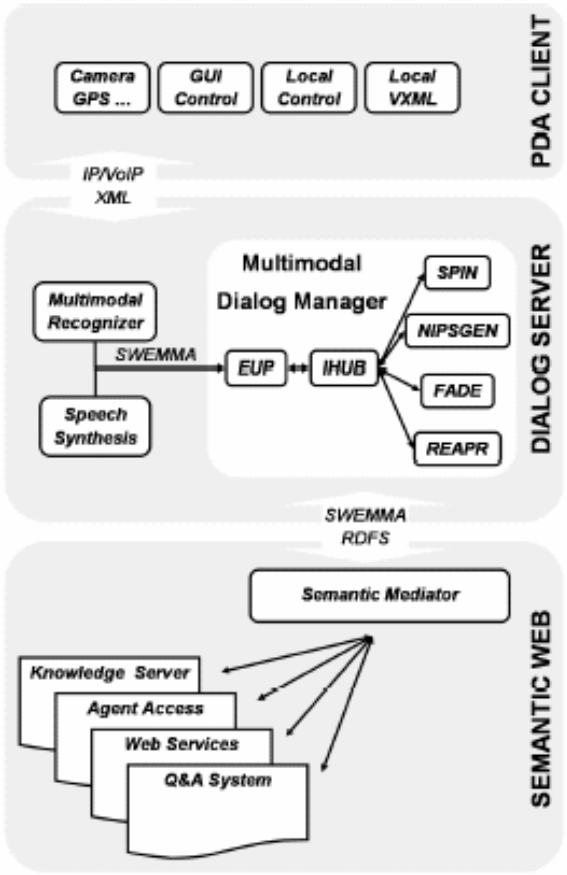
Integration of verbal and non-verbal output

Web Service Interface & System clarifications

Handheld Architecture

Implicit confirmation and language (in)dependent visualisation

Graphical Design



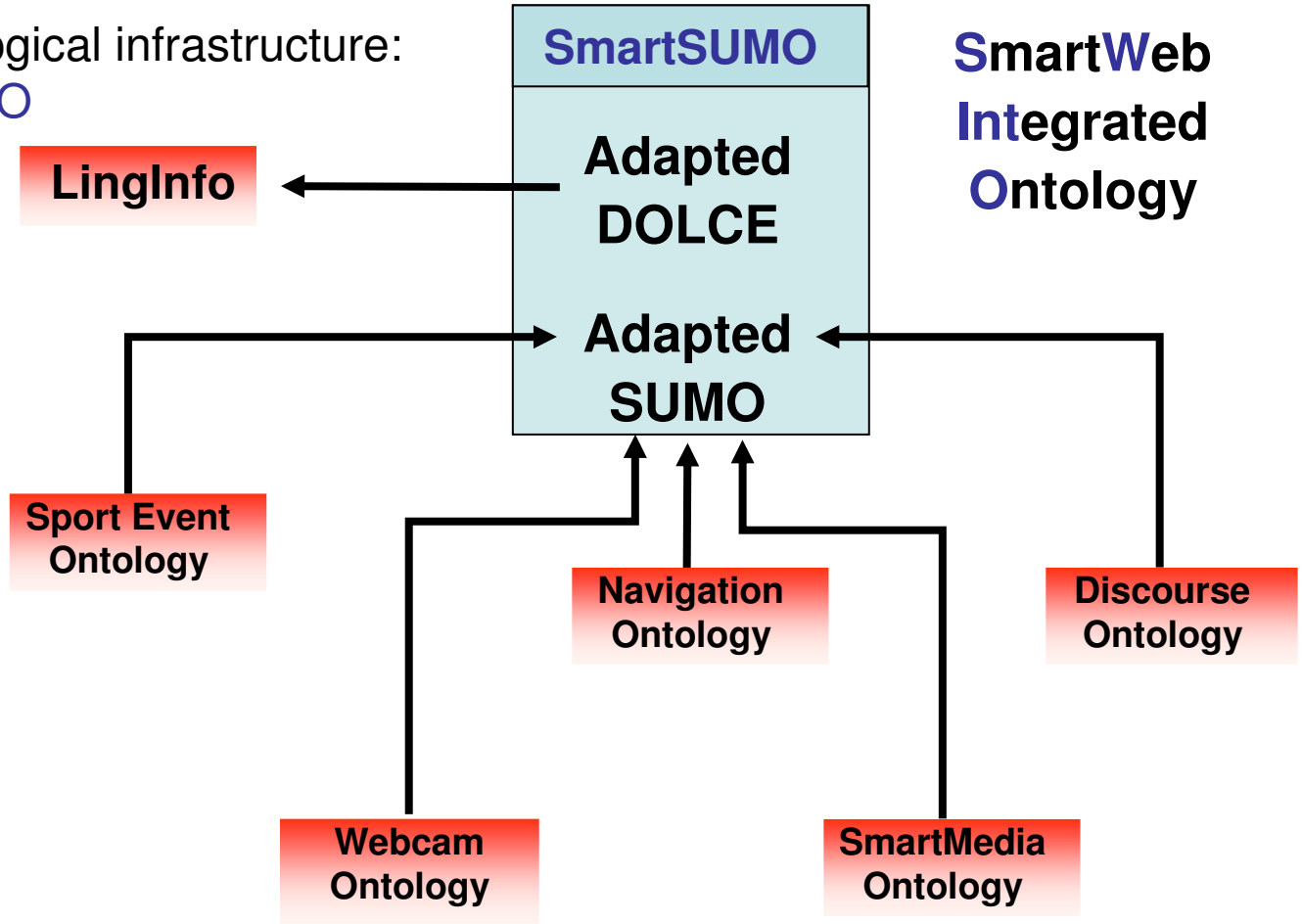
Technical Design

Ontologies

- An Ontology is
  - an explicit specification of a conceptualization [Gruber 93]
  - a shared understanding of a domain of interest [Uschold/Gruninger 96]
- Make domain assumptions **explicit**
  - Separate **domain knowledge** from operational knowledge
    - Re-use domain and operational knowledge separately
- A **community reference** for applications
- **Shared understanding** of what particular information means

Ontology Representation

Ontological infrastructure:
SWIntO



Discourse Ontology for Semantic Web Applications

Multimodal Dialogue Management

EMMA Format swemma:Result, swemma:Status

Multimodal question/ answer types

Dialogue Modelling

Lexical rules for syntactic/ semantic mapping

Dialogue Acts

Dialogue Memory

System reactions

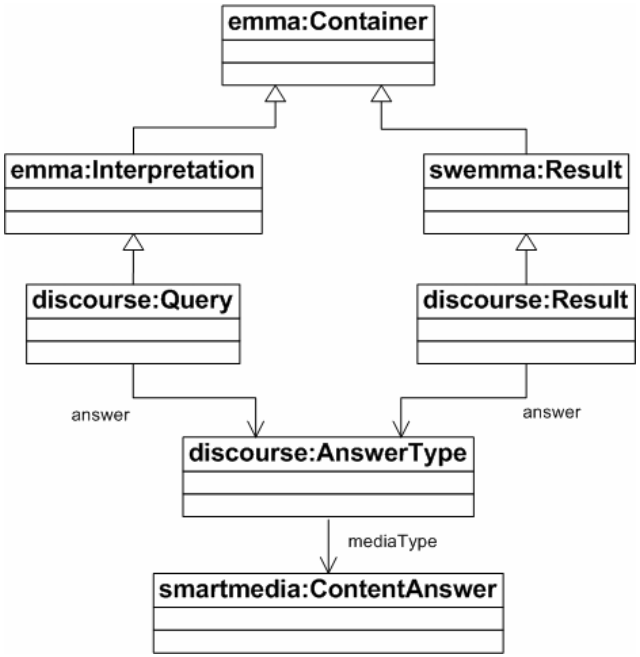
usage- and clarification dialogs
Presentation elements/patterns for multimodal fusion/fission

Dialogue Model

Phases/States, …
(FSM/PLAN)

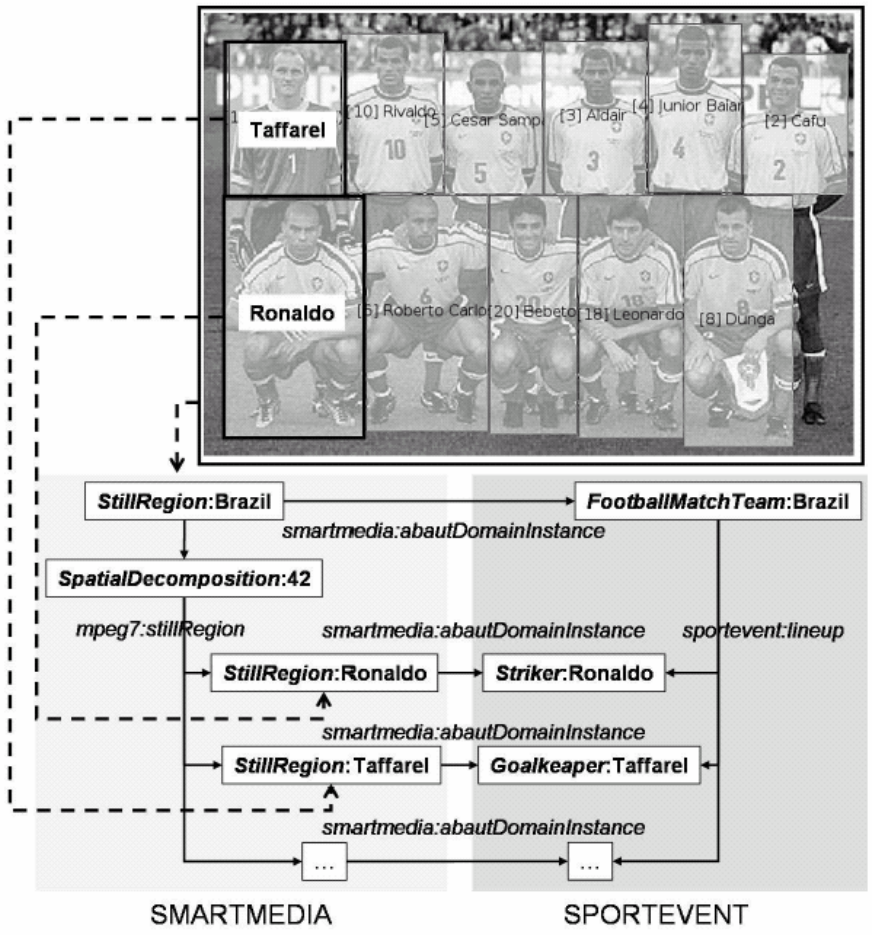
HCI

Interaction Patterns



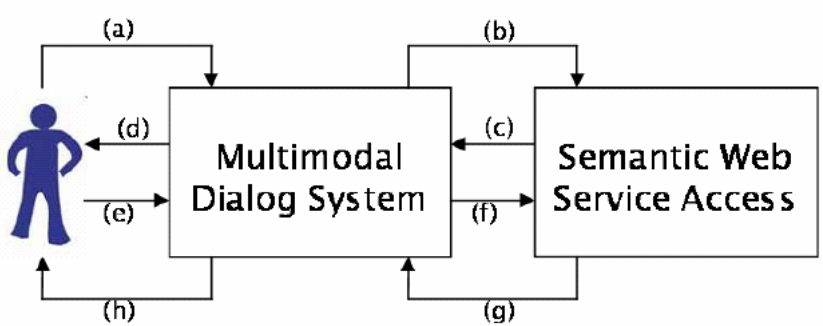
Ontology Representation and Multimedia

- Framework for gesture and speech fusion.
- Multimedia decomposition in space, time and frequency.
- Link to the Upper Model Ontology to close the *Semantic Gap.*



Ontology Representation and Web Services

- Detect underspecified user queries which lack required input parameters (GPS Missing).
- Plan-based Service Composer GOAL



(a) User query: What can I do in my spare time on Saturday?
(b) Ontological user query is sent to web services.
(c) Clarification request (asking for a city) is sent back.
(d) Verbalized clarification request: Where?
(e) User clarification response: In Berlin.
(f) Completed ontological query is sent to web services.
(g) Ontological result of service execution is sent to dialog.
(h) Generated results are multimodally presented to the user.

T-Traffic Route Planner, POI Search

info Movies, Events, Maps, Weather

SMARTWEB U-Context, e.g. p/t, Weather

amazon.de Books, Movie Posters

WM-Guide Train Conncetions, Hotels

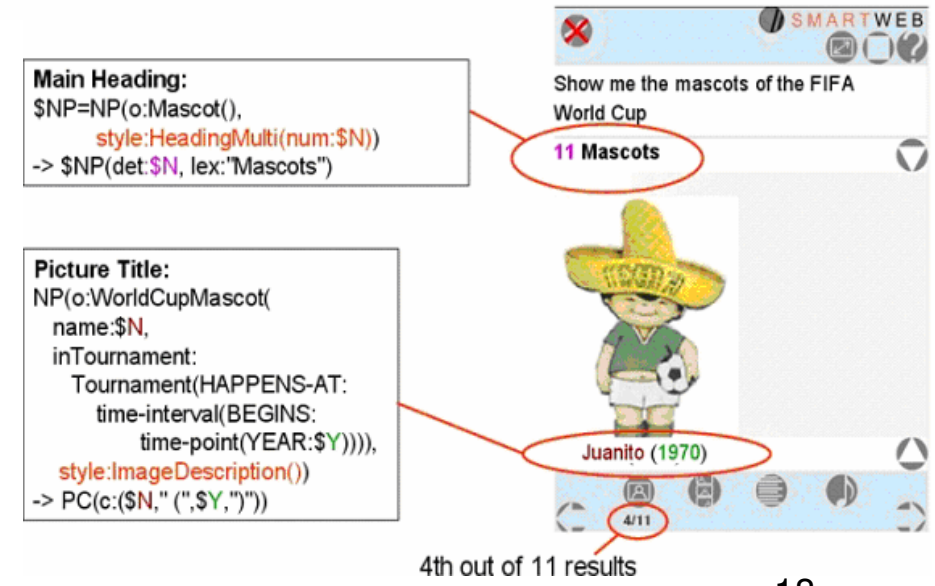
Web Service Composition Results

- Text-based event details, additional image material, and the location map are semantically represented.
- User perceptual feedback:
  - Feedback on natural language understanding
  - Presentation of Multimodal results combinjng text, image and speech synthesis

Language Understanding and Text Generation

- Lexico-Semantic Mapping on word level (SPIN).
- Fast and robust speech processing (ASR errors and disfluencies).
- Order-independent matching
- NIPSGEN uses SPIN + TAG grammar

```
[ discourse#Query
      text: "wer war mehr als zweimal Weltmeister"
      dialogueAct: [ InterrogQuestion ]
      focus: [ Focus
         focusMediumType:[ mpeg7#Text ]
         ...
         contextObject: [ FIFAWorldCup
                          winner:Team
                          origin: [ sumo#Country  ... ]
                        ]
         contextObject: [ GreaterThan
                          constraintRightArg: "2"
                        ]
         varName: ?X
         ]
]
```



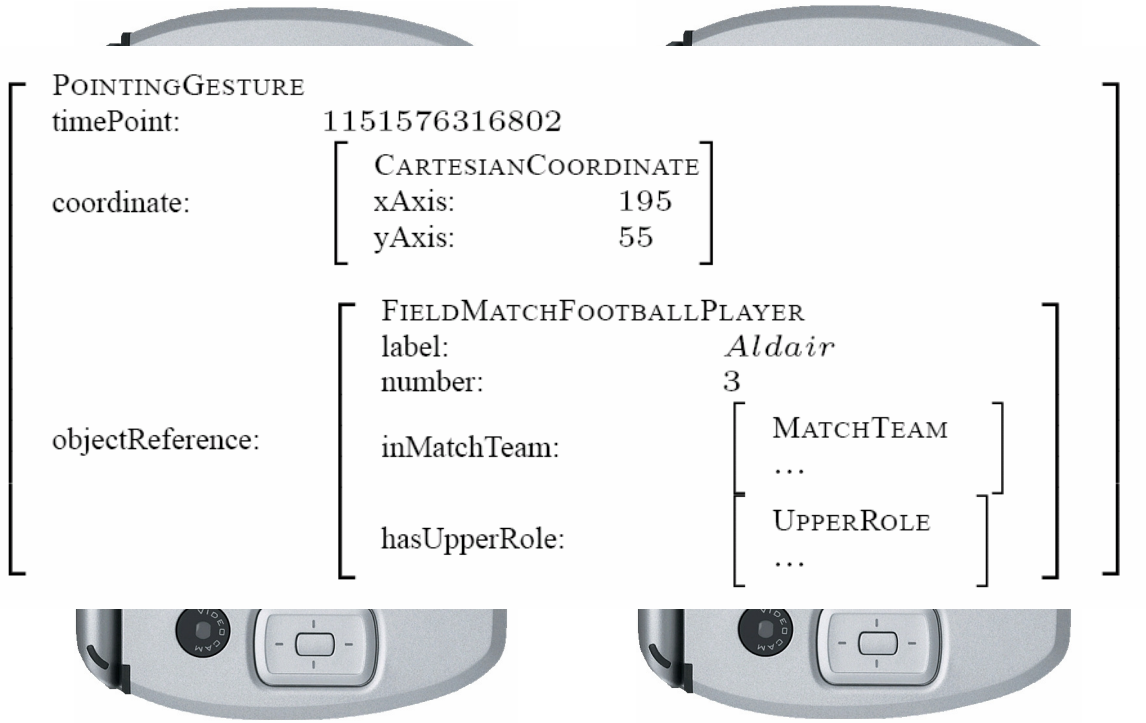
Multimodal Discourse Processing (FADE)

- Production rule system and discourse modeller
  - *"How do I get to Berlin from here?"*
  - *"Who won the Fifa World Cup in 1990?"*
  - *"And in 2002?"*
  - *"How often did this team [pointing gesture] win the World Cup?"*
  - *"What's the weather going to be like tomorrow?"*
  - *"And the day after tomorrow?"*
  - *"What's the name of the third player in the top row?"*.
  - *"Where do you want to start?"* - *"Berlin"*

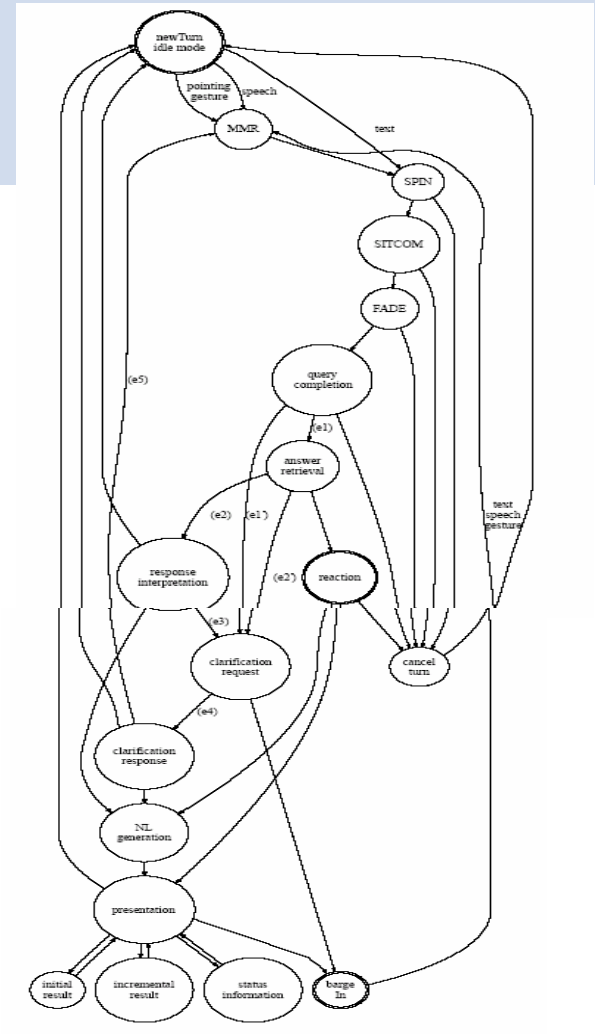
Multimodal Fusion

- Referencing Mpeg-7 sub-annotations:

[player click] + „How many goals did this player score.“



Multimodal Dialogue Control



- Initial FSA policy for demonstrator
- Outgoing arcs provide classification task
- IS features from ontological instances

- Learn Reaction and Presentation Behaviour
  - Clarification strategies
  - Filtering strategies
  - Multimodal presentation strategies with temporal constraints
  - User Modelling
  - Recommendations

Experience on Component Integration

- Using ontologies in information gathering dialog systems for knowledge retrieval from ontologies and web services in combination with advanced dialogical interaction is an iterative ontology engineering process.
- Ontological representations offer framework for gesture and speech fusion when users interact with Semantic Web results such as MPEG7-annotated images and maps.
- Generate structured input spaces for more context-relevant reaction planning to ensure naturalness in system-user interactions to a large degree.

Conclusions

- SmartWeb was sucessfully demonstrated in the context of the FootballWorld Cup 2006 in Germany.
- Flexible control flow to be combined with dialog system strategies for
  - error recoveries
  - clarifications with the user
  - multimodal interactions
- Future integration plans:
  - Dialogue management adaptations via machine learning
  - Collaborative filtering (of redundant results)
  - Incremental presentation of results