Federal Ministry
of Education
and Research

Daniel Sonntag, Paul Buitelaar

DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany

Overview

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- SmartWeb – Consortium and Goals
- Ontology-based Dialog
- Ontology-based Information Extraction
- Ontology Learning

The SmartWeb Consortium

SMARTWEB

Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

Goal: Context-Aware, Mobile, Multimodal Interface to the Semantic Web

SMARTWEB

Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- Multimodal coherent access to
  - Semantic Web services
  - Web based Question-Answering
  - Knowledge server
  - Agent based web wrappers
- Currently focused domain: A visit to a FIFA Football World Cup match in 2006
- Information about
  - Events and results
  - Sights
  - General events
- Towards open domain questions

Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

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.

Special Requirements

- Develop a context-aware, mobile, multimodal user interface.
- Use a smartphone as interaction device.
- Query transmission via UMTS, WLAN to the backend system.
- Barge-in, multiuser, permanent user control.
- Clear, ontology-based interface between modules (Jena API).
- Real-time interactive editing of semantic queries.

No off-the-shelf solution

Ontology-based Dialog

Norbert Reithinger, Simon Bergweiler, Ralf Engel, Gerd Herzog,
Norbert Pfleger, Massimo Romanelli, Daniel Sonntag

Devices and Modalities

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

Interaction devices

- 3G Smartphone
- Car
- Motor-bike

Modalities:

- Speech
- Gestures
- Haptic interaction
- Face camera
- Bio-signals

Interaction Example

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- 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 …

Interaction Design: Requirements

SMARTWEB
Funded by Federal Ministry of Education and Research
Grant 01 IMD01

- User actions:
  - Ask simple factoid and inspection questions or commands (*"Who won the 1978 championship?"*)
  - Search, explore, and inspect information
  - Control the system

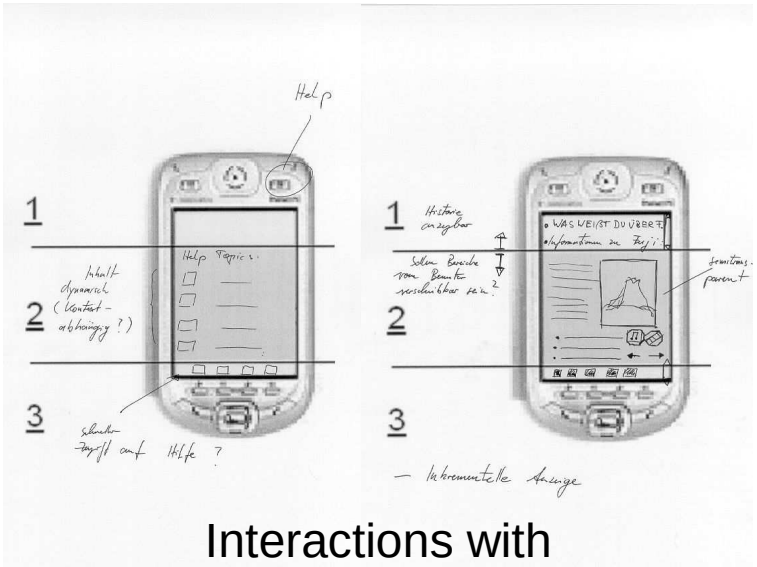
- Interaction design principles:
  - Follow the UI principles of the mobile device
  - Display the recognized user input
  - Offer correction possibilities
  - Provide interface simplicity by progressive disclosure
  - Provide status information to the user

From Storyboard to Realisation

SMARTWEB
Funded by Federal Ministry of Education and Research
Grant 01 IMD01

- Three interaction areas
  - User input region
  - Result presentation
  - Status and control icons/buttons

Initial screen layout



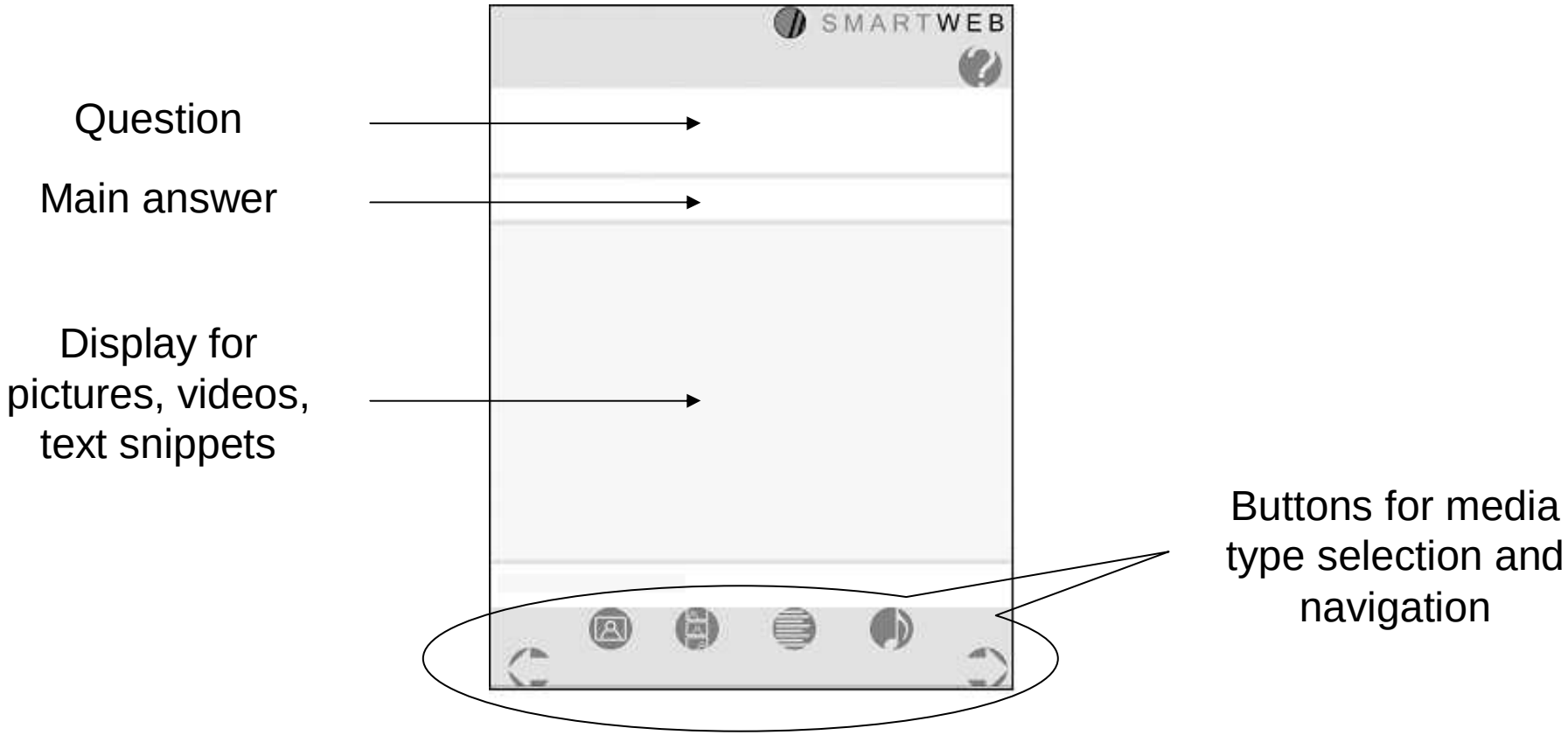
Interactions with paper prototypes

Implemented system

Interface Design Elements

SMARTWEB
Funded by Federal Ministry of Education and Research
Grant 01 IMD01

Design adapted to screen size and interaction metaphor

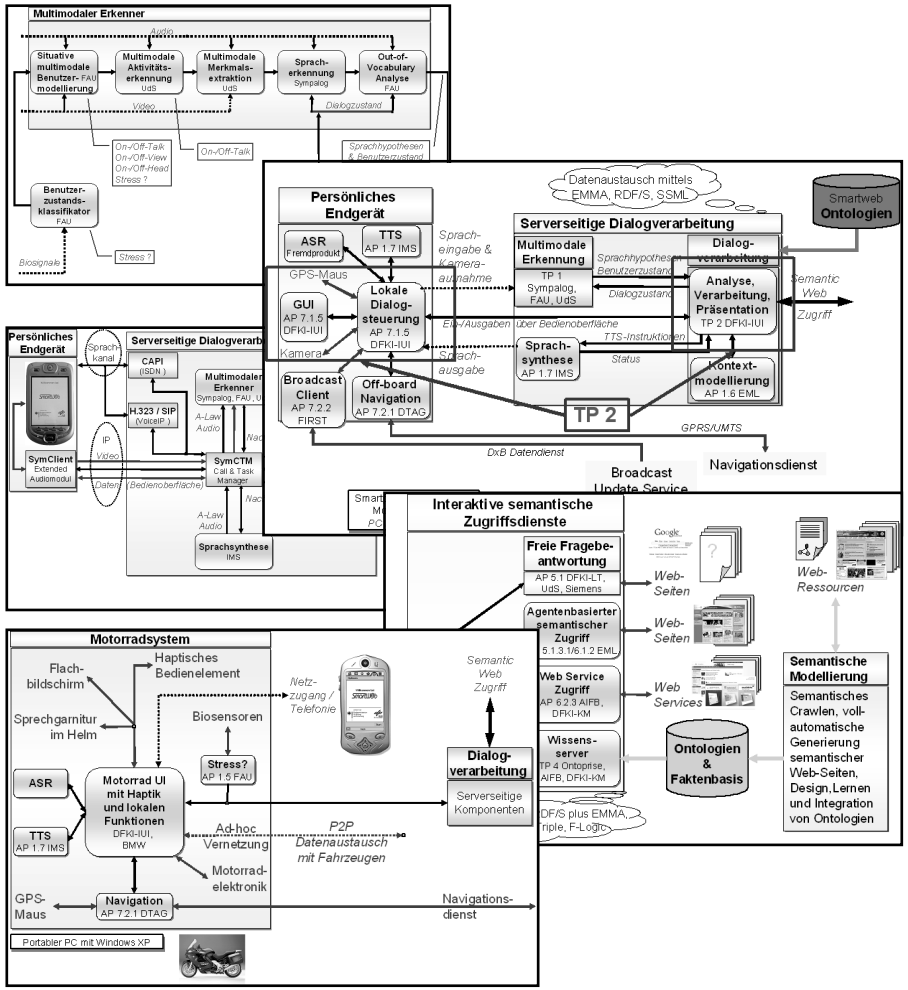


Realized in Macromedia Flash, local processing in Actionscript

Principles of System Design

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

Controlling the complexity:

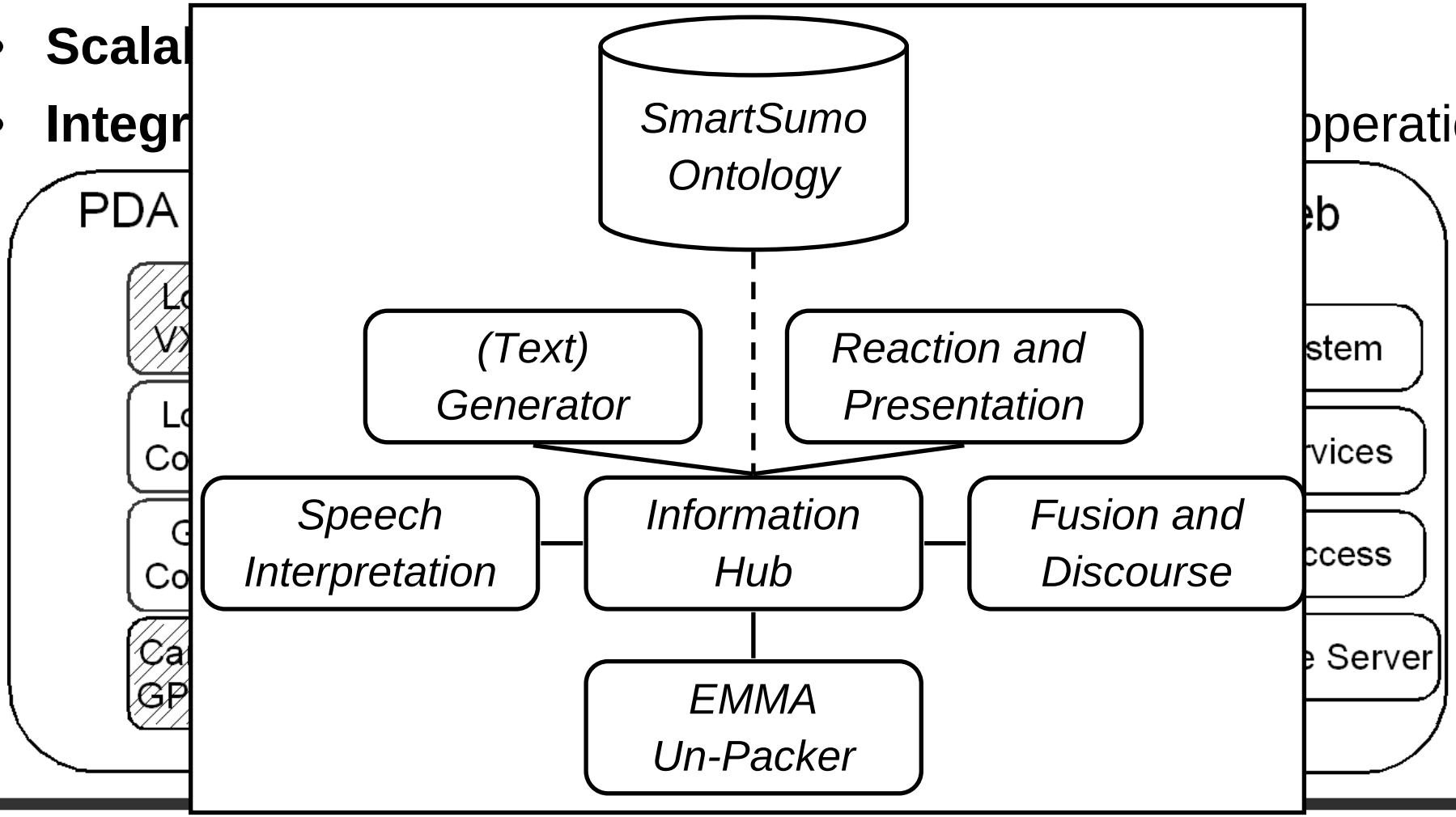


- **Representation**
  - provides a common ground of terms and structures
  - is part of the global dialog history in the discourse module
- **Standards**
  - ease scalability
  - enable re-use
- **Interfaces**
  - define representation for module communication
- **Encapsulation**
  - separates the interface from application logic

Funded by
Federal Ministry
of Education
and Research
Grant 01 IMD01

General Architecture

- **Openness**, especially for Semantic Web technology (ontology driven processing)
- **Scalal**
- **Integr** operations



Using EMMA

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- Infor…dal Ann…
  - F… n…
  - G…
  - W…
  - S…ed
- Exte…
  - **r…**
  - **s…** p…
  - **o…**
- RDF…e onto…

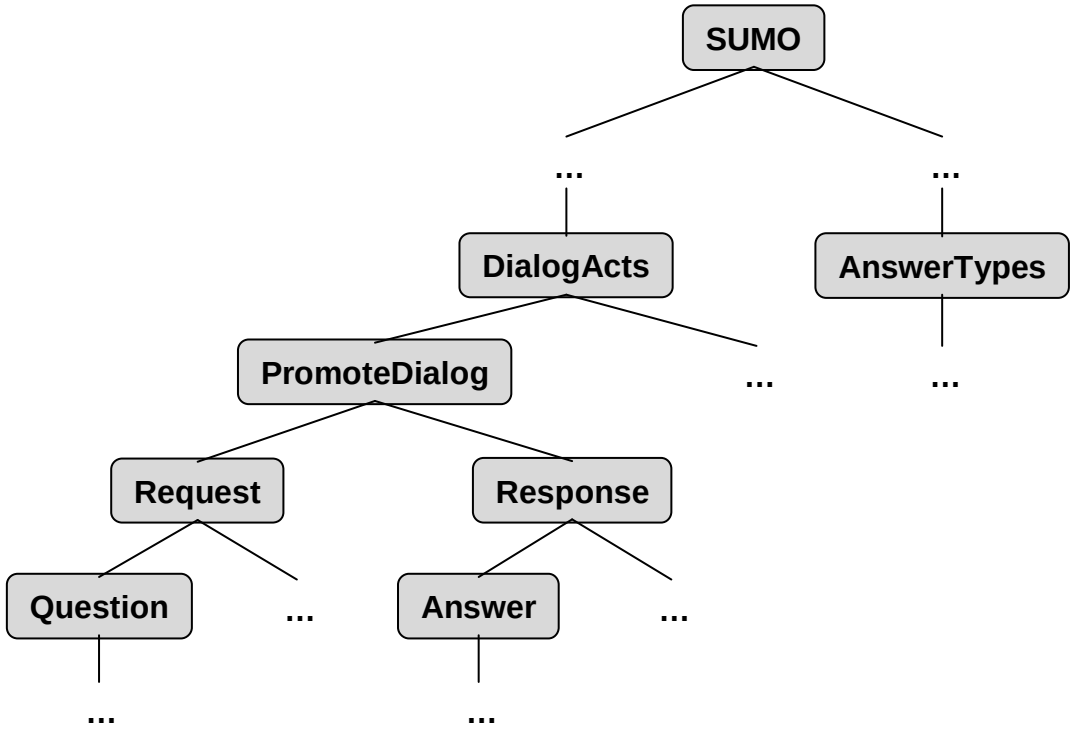
SWEMMA encoded output for the speech synthesizer:

```
<emma:emma version="1.0">
    <swemma:result emma:id="DIA123"
                   emma:process="DIA#42"
                   emma:turn-id="42"
                   emma:lang="de“
                   emma:start="1087995961542"
                   emma:end="1087995963542"
                   emma:mode="speech">
        <swemma:result emma:confidence="1.0">
            <emma:derived-from emma:resource="#spin1"/>
            <speak version="1.0“  xsi:schemaLocation="www.w3c.org/....xsd"
                   xml:lang="de">
                1990 war Deutschland Fussballweltmeister.
            </speak>
        </swemma:result>
    </swemma:result>
</emma:emma>
```

Common Ontology

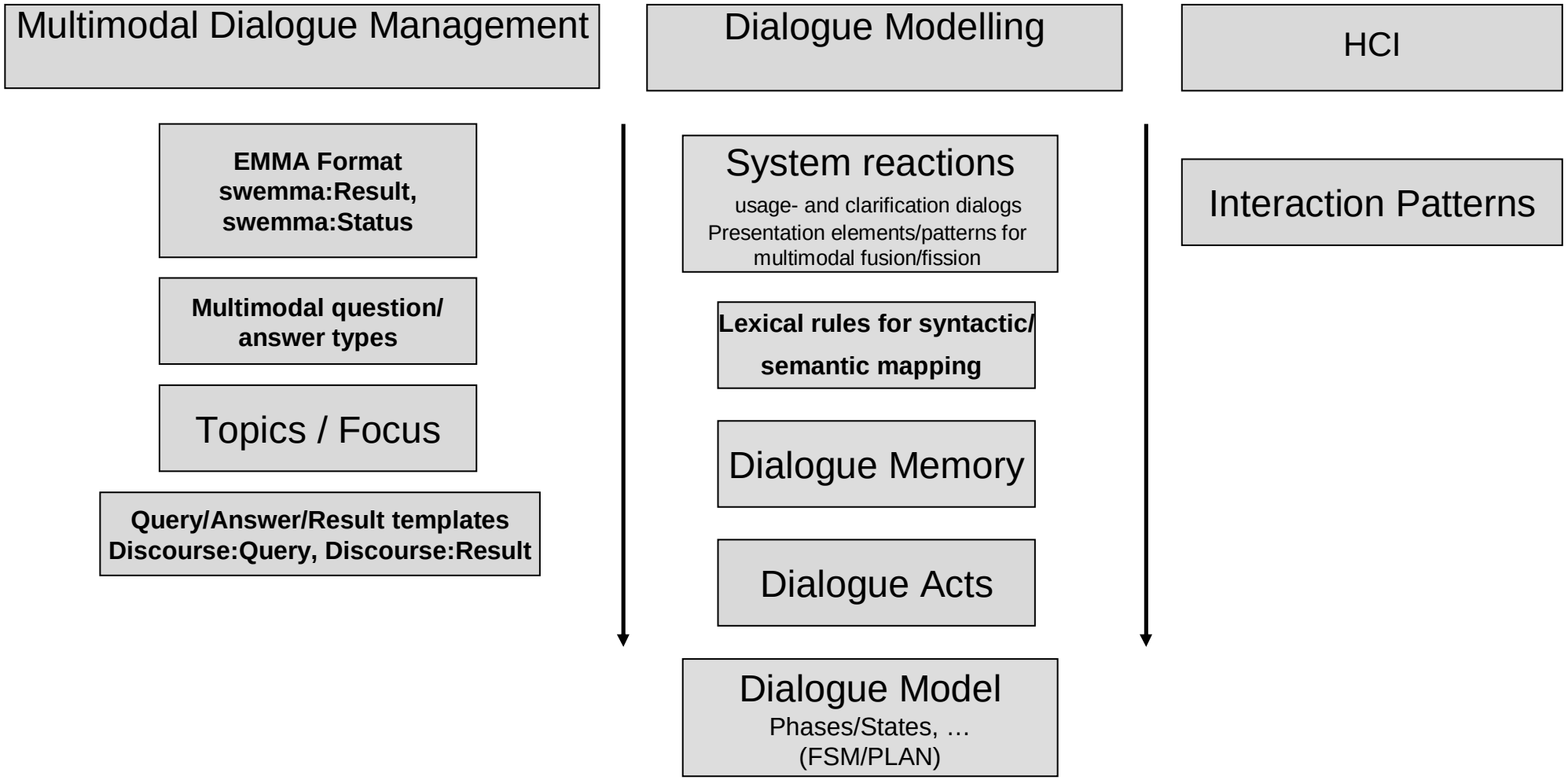
SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- SmartSumo ontology covers
  - General knowledge base (SUMO+DOLCE)
  - Domain specific knowledge: sports events, navigation, …
  - Requests for web services
  - Control commands
- Discourse ontology
  - **`Question`**: user input
  - **`Answer`**: processing result, possibly multiple answer type instances
  - **`MediumType`**: description of the resulting media



Discourse Ontology for Semantic Web applications

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01



```
<rdf:RDF
  xmlns:jms="http://jena.hpl.hp.com/2003/08/jms#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:j.0="http://smartweb.org/ontology/emma#"
  xmlns:j.1="http://smartweb.org/ontology/discourse#"
  xmlns:j.2="http://smartweb.org/ontology/sportevent#">
  <j.0:Emma>
    ...
    <j.1:Query rdf:about="http://smartweb.org/ind#i4">
      <j.1:text rdf:datatype=
        "http://www.w3.org/2001/2001/XMLSchema#string"
        >wer war 1990 Weltmeister</j.1:text>
      <j.1:dialogueAct>
        <j.1:Question />
      </j.1:dialogueAct>
      <j.1:focus>
        <j.2:DivisionNationalTeam
          rdf:about="http://smartweb.org/ind#i5"/>
      </j.1:focus>
      <j.1:content>
        <j.2:WorldCup>
          <j.2:heldOn rdf:datatype=
            "http://www.w3.org/2001/XMLSchema#string"
            >1990</j.2:heldOn>
          <j.2:winner
            rdf:resource="http://smartweb.org/ind#i5"/>
        </j.2:WorldCup>
      </j.1:content>
      <j.0:confidence rdf:datatype=
        "http://www.w3.org/2001/XMLSchema#float"
        >0.75</j.0:confidence>
    </j.1:Query>
    ...
  </j.0:Emma>
</rdf:RDF>
```

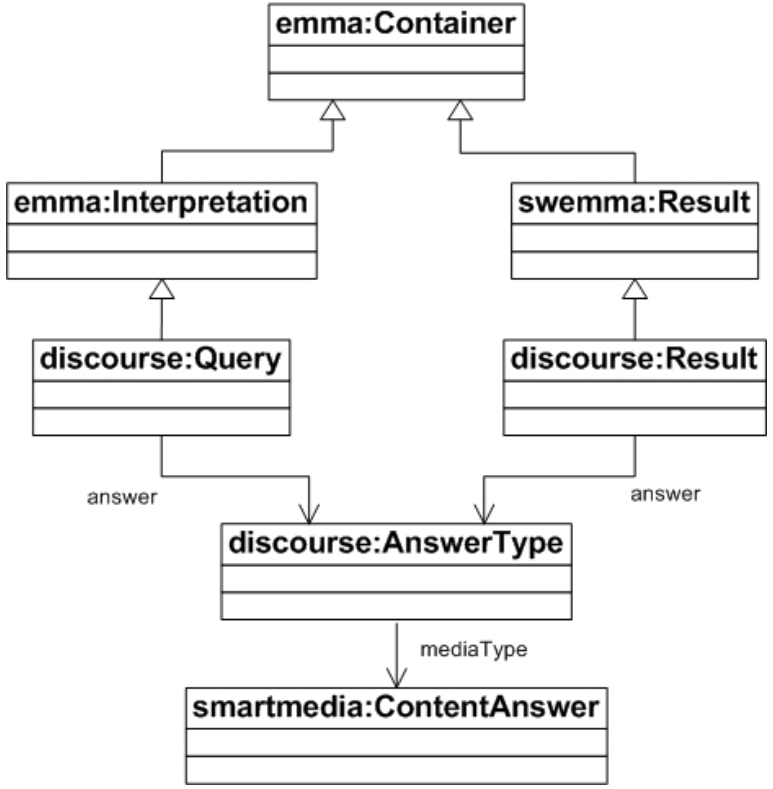


Figure 9: Example for the use of the Query concept in the interpretation of the utterance *"wer war 1990 Weltmeister?"* (*who was world champion in 1990?*). The tag `j2:content` contains a partially filled ontology instance to be completed by result information.

Current State: Demonstrator System V 0.1

SMARTWEB

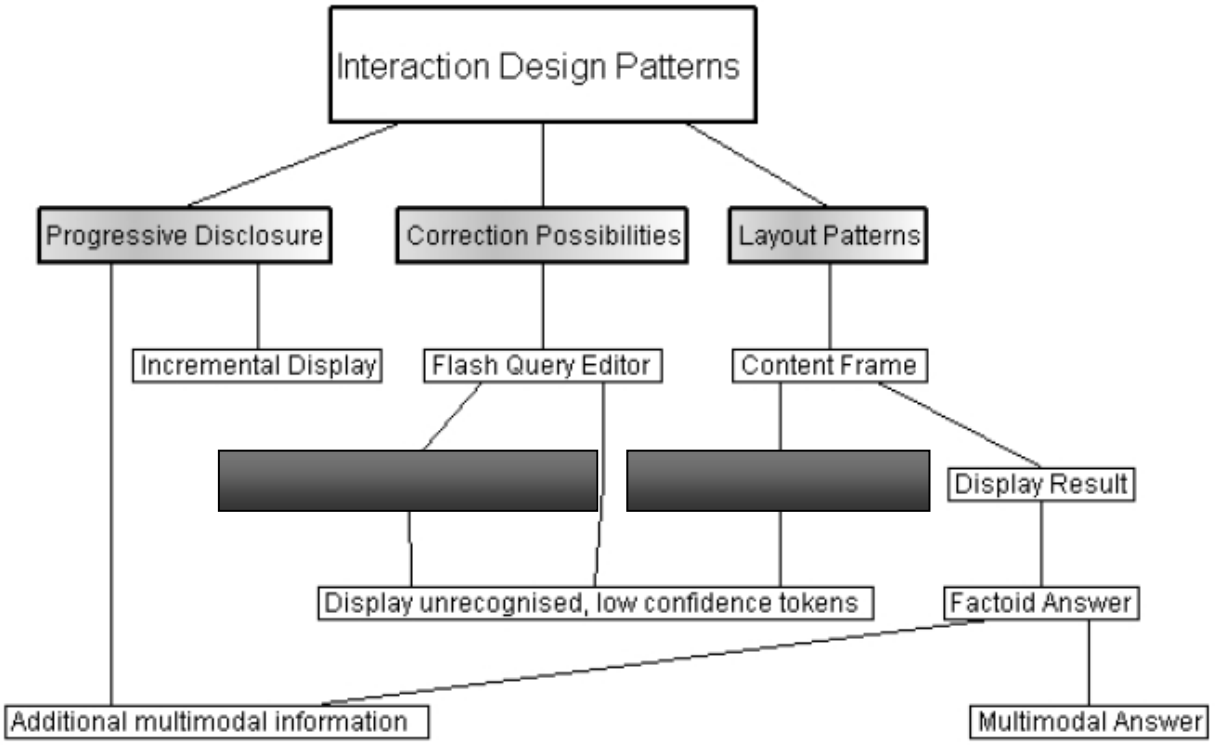
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

Interaction Design Patterns

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

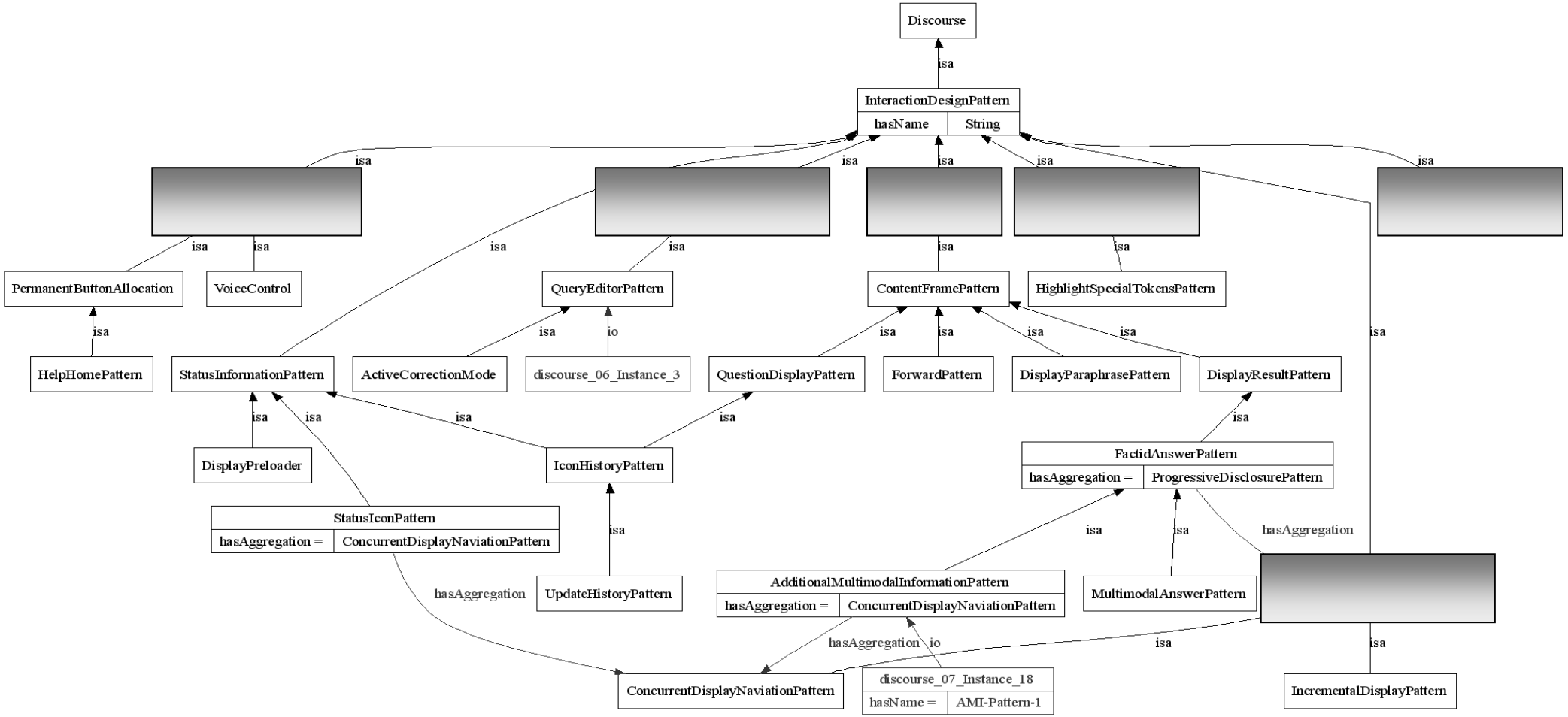
```
User_Correction -> {More_active_correction_mode,
Display_unrecognised_tokens}.

User_Query ->  {Display_Paraphrase}.
```



Interaction ontology

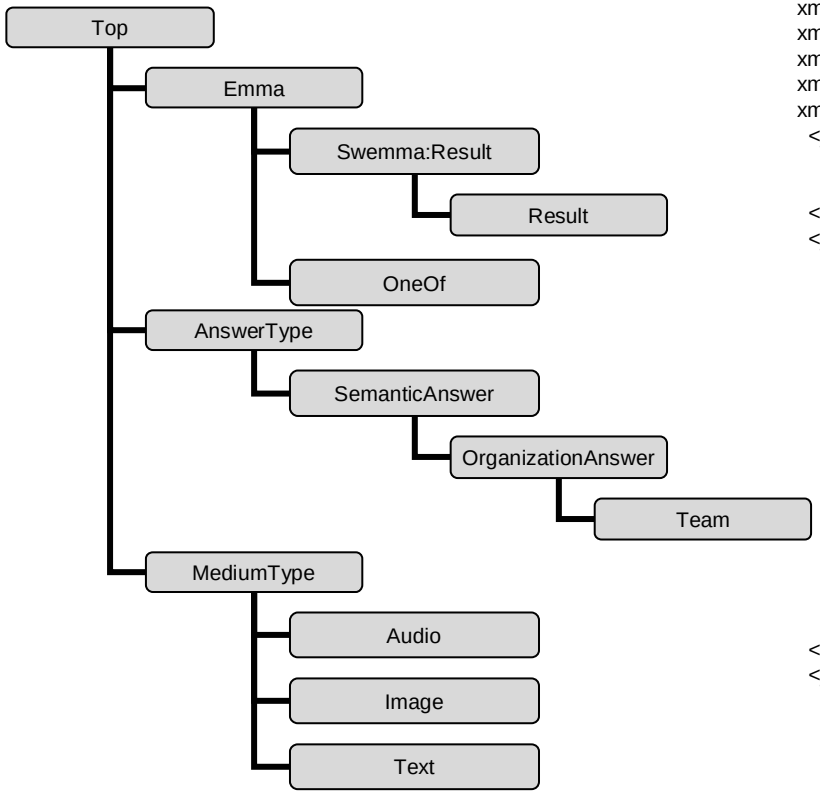
SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01



Result Representation: *Who was world champion in 1990?*

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

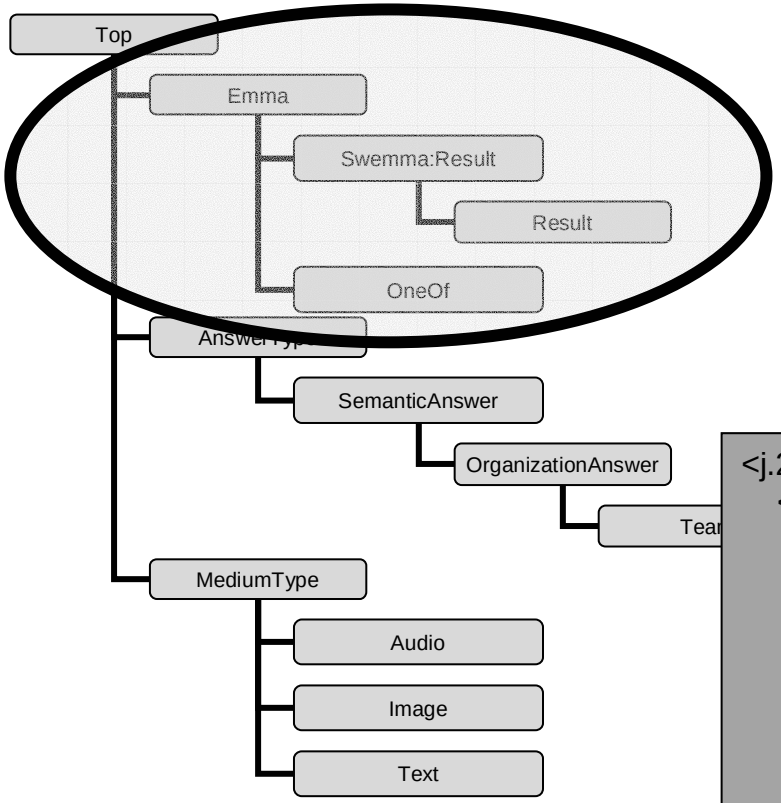
Type hierarchy:




```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:j.0="http:// ... /swemma#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:j.1="http:// ... /discourse#"
xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:j.2="http:// ... /emma#">
 <j.1:Text rdf:about="http:// ... /discourse#reapr_text_instance_Weltmeister1990">
  <j.1:sourceURI>http://fifaworldcup.yahoo.com/06/de/o/octeam/emb.html</j.1:sourceURI>
  <j.1:Content>DEUTSCHLAND</j.1:Content>
 </j.1:Text>
 <j.1:Result rdf:about="http:// ... /discourse#result_instance_Weltmeister1990">
  <j.1:answerTypes>
   <j.1:Team rdf:about="http:// ... /discourse#reapr_team_Weltmeister1990">
    <j.1:mediaTypes rdf:resource="http:// ... /discourse#reapr_text_instance_Weltmeister1990"/>
    <j.1:mediaTypes>
     <j.1:Image rdf:about="http:// ... /discourse#reapr_image_instance_Weltmeister1990">
      <j.1:sourceURI> http://www.zdf.de/ZDFde/img/28/0,1886,2425724,00.jpg</j.1:sourceURI>
     </j.1:Image>
    </j.1:mediaTypes>
    <j.1:mediaTypes>
     <j.1:Text rdf:about="http:// ... /discourse#reapr_text2_instance_Weltmeister1990">
      <j.1:Content>GERMANY, I want beer.</j.1:Content>
     </j.1:Text>
    </j.1:mediaTypes>
   </j.1:Team>
  </j.1:answerTypes>
 </j.1:Result>
 <j.2:Emma rdf:about="http:// ... /emma#emma_root_instance">
  <j.2:container>
   <j.0:Result rdf:about="http:// ... /swemma#reapr_swemma_result_instance_Weltmeister1990">
    <j.2:container>
     <j.2:OneOf rdf:about="http:// ... /emma#reapr_oneOf_instance_1">
      <j.2:container rdf:resource="http:// ... /discourse#result_instance_Weltmeister1990"/>
     </j.2:OneOf>
    </j.2:container>
   </j.0:Result>
  </j.2:container>
  <j.2:version>1.0</j.2:version>
 </j.2:Emma>
</rdf:RDF>
```

Result Representation:
Who was world champion in 1990?

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01



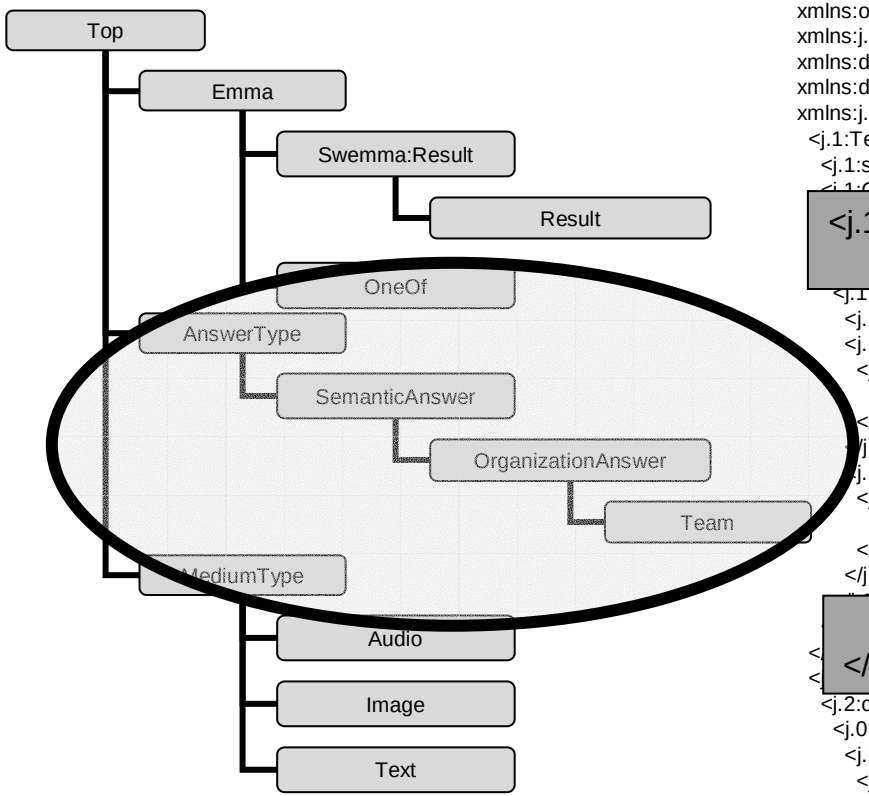

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:j.0="http:// ... /swemma#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:j.1="http:// ... /discourse#"
xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:j.2="http:// ... /emma#">
 <j.1:Text rdf:about="http:// ... /discourse#reapr_text_instance_Weltmeister1990">
  <j.1:sourceURI>http://fifaworldcup.yahoo.com/06/de/o/octeam/emb.html</j.1:sourceURI>
  <j.1:Content>DEUTSCHLAND</j.1:Content>
 </j.1:Text>
 <j.1:Result rdf:about="http:// ... /discourse#result_instance_Weltmeister1990">
  <j.1:answerTypes>
   <j.1:Team rdf:about="http:// ... /discourse#reapr_team_Weltmeister1990">
    <j.1:mediaTypes rdf:resource="http:// ... /discourse#reapr_text_instance_Weltmeister1990"/>
    <j.1:mediaTypes>
     <j.1:Image rdf:about="http:// ... /discourse#reapr_image_instance_Weltmeister1990">
      <j.1:sourceURI> http://www.zdf.de/ZDFde/img/28/0,1886,2425724,00.jpg</j.1:sourceURI>
     </j.1:Image>
```

```
<j.2:Emma rdf:about="http:// … /emma#emma_root_instance">
  <j.2:container>
   <j.0:Result rdf:about="http:// … /swemma#reapr_swemma_result_instance_Wm1990">
    <j.2:container>
     <j.2:OneOf rdf:about="http:// ... /emma#reapr_oneOf_instance_1">
      <j.2:container rdf:resource="http:// … /discourse#result_instance_Wm1990"/>
     </j.2:OneOf>
    </j.2:container>
   </j.0:Result>
  </j.2:container>
  <j.2:version>1.0</j.2:version>
 </j.2:Emma>
```

```
    </j.2:Emma>
</rdf:RDF>
```

Result Representation:
Who was world champion in 1990?

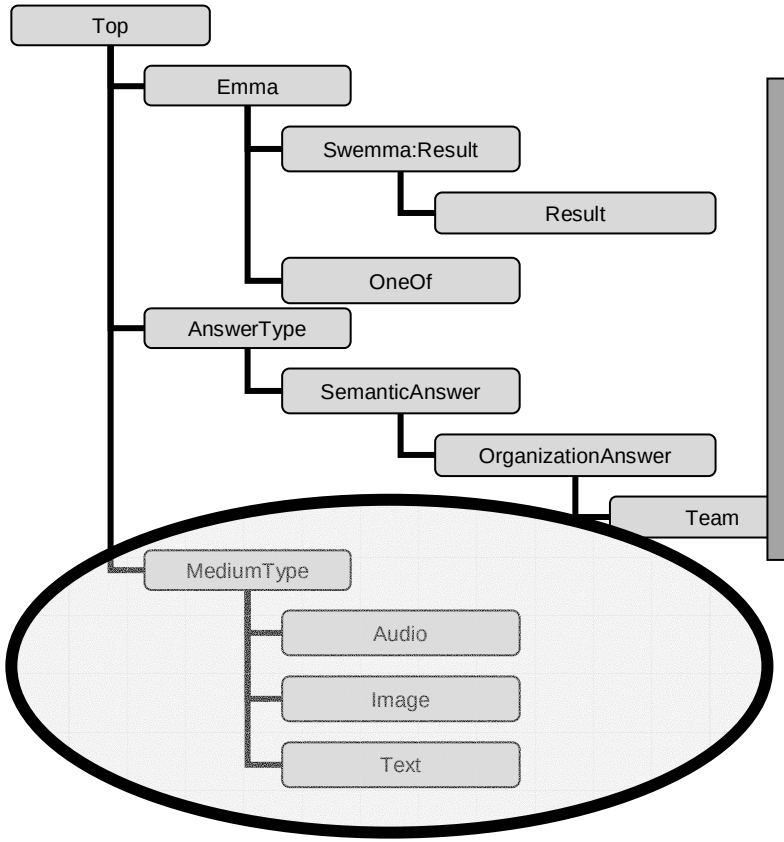
SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01




```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:j.0="http:// ... /swemma#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:j.1="http:// ... /discourse#"
xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:j.2="http:// ... /emma#">
 <j.1:Text rdf:about="http:// ... /discourse#reapr_text_instance_Weltmeister1990">
  <j.1:sourceURI>http://fifaworldcup.yahoo.com/06/de/o/octeam/emb.html</j.1:sourceURI>
  <j.1:Content>DEUTSCHLAND</j.1:Content>
  <j.1:answerTypes>
    <j.1:Team rdf:about="http:// ... /discourse#reapr_team_Wm1990">
  <j.1:Team rdf:about="http:// ... /discourse#reapr_team_Weltmeister1990">
   <j.1:mediaTypes rdf:resource="http:// ... /discourse#reapr_text_instance_Weltmeister1990"/>
   <j.1:mediaTypes>
    <j.1:Image rdf:about="http:// ... /discourse#reapr_image_instance_Weltmeister1990">
     <j.1:sourceURI> http://www.zdf.de/ZDFde/img/28/0,1886,2425724,00.jpg</j.1:sourceURI>
    </j.1:Image>
   </j.1:mediaTypes>
   <j.1:mediaTypes>
    <j.1:Text rdf:about="http:// ... /discourse#reapr_text2_instance_Weltmeister1990">
     <j.1:Content>GERMANY, I want beer.</j.1:Content>
    </j.1:Text>
   </j.1:mediaTypes>
    </j.1:Team>
  </j.1:answerTypes>
 <j.2:container>
  <j.0:Result rdf:about="http:// ... /swemma#reapr_swemma_result_instance_Weltmeister1990">
   <j.2:container>
    <j.2:OneOf rdf:about="http:// ... /emma#reapr_oneOf_instance_1">
     <j.2:container rdf:resource="http:// ... /discourse#result_instance_Weltmeister1990"/>
    </j.2:OneOf>
   </j.2:container>
  </j.0:Result>
 </j.2:container>
 <j.2:version>1.0</j.2:version>
</j.2:Emma>
</rdf:RDF>
```

Result Representation:
Who was world champion in 1990?

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01




```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:j.0="http:// ... /swemma#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:j.1="http:// ... /discourse#"
xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
```

```
<j.1:mediaTypes rdf:resource="http:// … /discourse#reapr_text_instance_Wm1990"/>
<j.1:mediaTypes>
 <j.1:Image rdf:about="http:// … /discourse#reapr_image_instance_Wm1990">
  <j.1:sourceURI> http://..../img/28/0,1886,244,00.jpg </j.1:sourceURI>
 </j.1:Image>
</j.1:mediaTypes>
<j.1:mediaTypes>
 <j.1:Text rdf:about="http:// … /discourse#reapr_text2_instance_Wm1990">
  <j.1:Content>DÉUTSCHLAND.</j.1:Content>
 </j.1:Text>
</j.1:mediaTypes>
```

```
      </j.1:mediaTypes>
     </j.1:Team>
    </j.1:answerTypes>
   </j.1:Result>
   <j.2:Emma rdf:about="http:// ... /emma#emma_root_instance">
    <j.2:container>
     <j.0:Result rdf:about="http:// ... /swemma#reapr_swemma_result_instance_Weltmeister1990">
      <j.2:container>
       <j.2:OneOf rdf:about="http:// ... /emma#reapr_oneOf_instance_1">
        <j.2:container rdf:resource="http:// ... /discourse#result_instance_Weltmeister1990"/>
       </j.2:OneOf>
      </j.2:container>
     </j.0:Result>
    </j.2:container>
    <j.2:version>1.0</j.2:version>
   </j.2:Emma>
  </rdf:RDF>
```

SmartWeb Demonstrator System V 0.1

SMARTWEB

Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

Show me the championship's mascot!

How high is Mount Fuji?

Summary and Next Steps

SMARTWEB
Funded by Federal Ministry of Education and Research
Grant 01 IMD01

- √ Interaction design finished
- √ Flexible architecture realised
- √ Interface and content representations use and extend W3C standards like EMMA and RDF Schema
- √ First version is operable

- Next steps
  - User intervention/editing of speech input
  - Multi-user access
  - Extension of coverage

- Next versions at CeBit 2006 and at the FIFA Football World Cup 2006 in summer 2006

Ontology-based Information Extraction

Paul Buitelaar, Thomas Eigner, Greg Gulrajani, Alexander Schutz,
Melanie Siegel, Nicolas Weber (DFKI LT)

&

Philipp Cimiano (AIFB, Univ. of Karlsruhe)

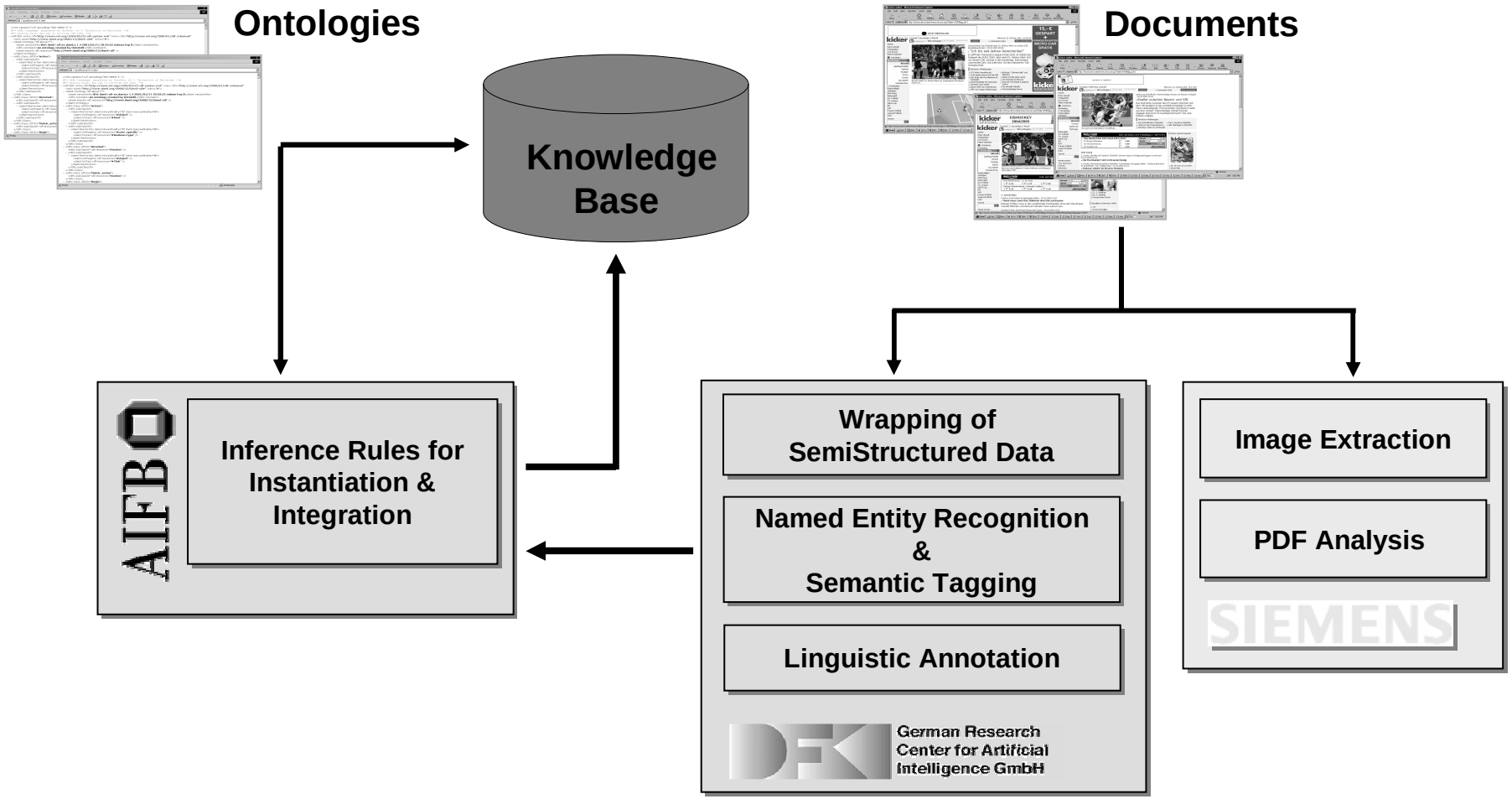
SOBIE *(SmartWeb Ontology-Based Information Extraction)*

SMARTWEB

Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- Ontology-Based Information Extraction …
  - Combines
    - Semantic Wrapping of Semi-Structured Data
    - Semantic and Linguistic Annotation of Free Text
    - Inference Rules for Instantiation and Integration of Annotated Entities and Events
- … and Display
  - Ontology-driven Hyperlink Generation for Display of Extracted Information
- Application Context
  - SmartWeb Project – Large, German Funded Project around World-Cup 2006
  - Intelligent, Mobile Information Services with Application Scenarios for the Pedestrian (Deutsche Telekom), Motorbike (BMW) and Car (DaimlerChrysler)
  - Integrates
    - Multimodal Dialog Processing
    - IR-based Question Answering
    - **Ontology-Based Information Extraction**
    - Semantic Web Services

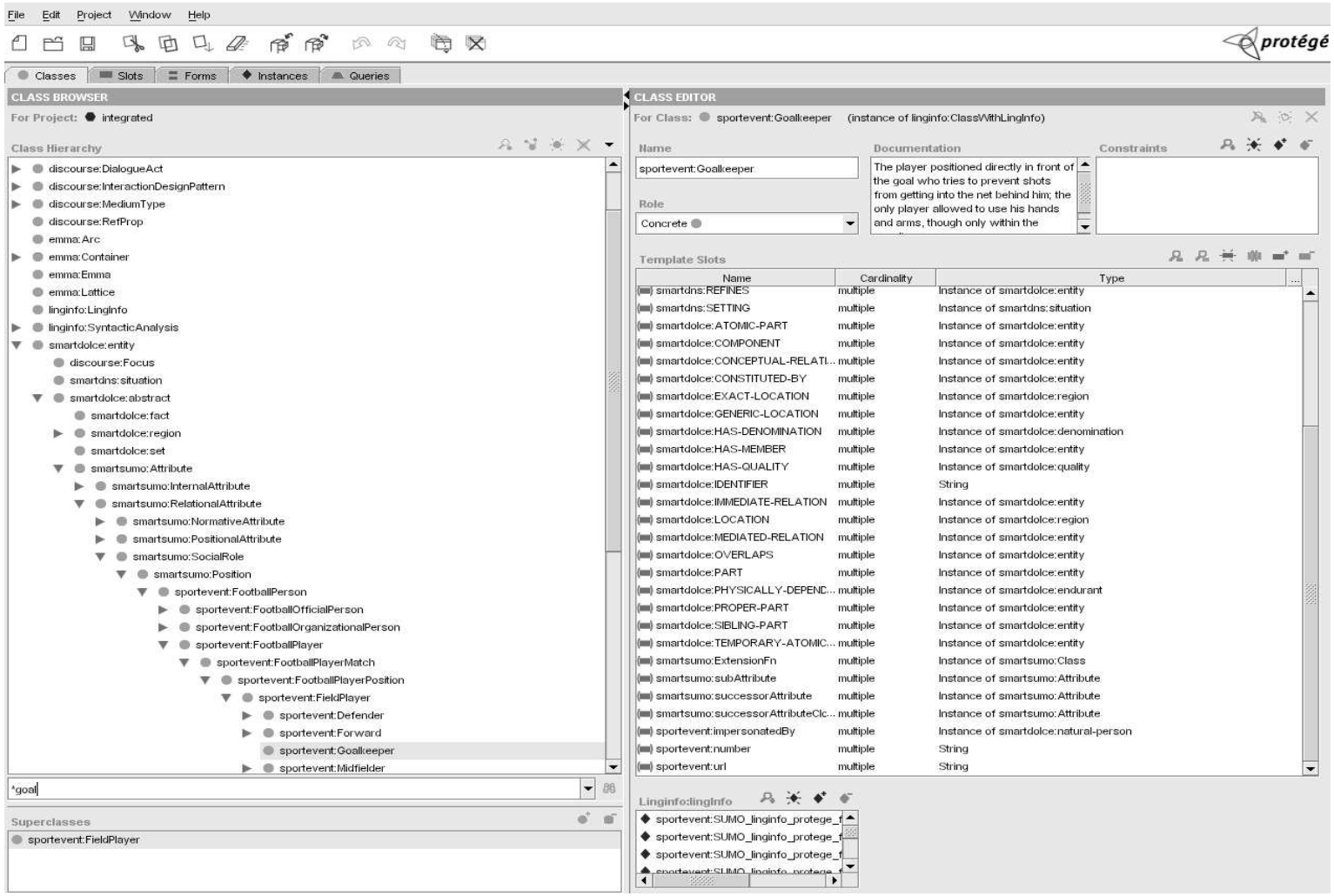
SOBIE – Processing and Data Flow



SmartWeb Integrated Ontology (by AIFB, DFKI KM/IUI, EML)

Funded by
Federal Ministry
of Education
and Research
Grant 01 IMD01

Integrates DOLCE (Foundational), SUMO (Top), SportEvents, Discourse and Navigation Ontology



SmartWeb Corpus

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- (Growing) Web Corpus through Monitor on
  - http://fifaworldcup.yahoo.com/
  - http://www.uefa.com/competitions/worldcup
- Semi-Structured Data
  - Tabular: Match Reports, Teams, etc.
- Free Text
  - Match Reports
  - Image Captions

Semi-Structured Data - HTML

Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

FIFAworldcup.com YAHOO! Fenster Schließen

Match Report

Fédération Internationale de Football Association
Hitzigweg 11 P.O. Box 85 8030 Zurich Switzerland Tel: 41-1/384 95 95 Fax: 41-1/384 96 96

FIFA For the Good of the Game

Bolivien - Kolumbien **1:1 (1:1)**

Spiel	Datum	Spielort	Stadion	Zeit	Spectators
	26-APR-00	La Paz	Hernando Siles	12:00	35500

Tore
SANCHEZ Erwin (BOL) 16', CASTILLO Jairo (COL) 32'.

Bolivien

[1]	FERNANDEZ Jose (GK)
[2]	PENA Juan
[3]	SANDY Marco (-32')
[6]	SORIA Vladimir
[7]	RIBEIRO Luis
[8]	CRISTALDO Luis
[9]	ANTELO Victor (-45')
[10]	GUTIERREZ Raul (-63')
[11]	MORENO Jaime
[19]	CASTILLO Ivan
[21]	SANCHEZ Erwin

Substitutes

[4]	RIMBA Miguel (+32')
[12]	SORIA Mauricio
[14]	BOTERO Joaquin
[16]	GALINDO Gonzalo (+63')
[17]	SUAREZ Roger (+45')
[18]	JUSTINIANO Miguel
[20]	RIBERA Reny

Trainer ARAGONES Carlos

Kolumbien

[1]	CORDOBA Oscar (GK)
[2]	CORDOBA Ivan
[3]	YEPES Mario
[4]	VIVEROS Alexander
[5]	BERMUDEZ Jorge
[6]	DINAS Arley
[8]	OVIEDO Franki (-81')
[11]	CASTILLO Jairo
[15]	MARTINEZ Gonzalo (-79')
[18]	RICARD Hamilton (-65')
[19]	RINCON Freddy

Substitutes

[7]	CASTRO Carlos
[9]	ANGEL Juan Pablo (+65')
[12]	CALERO Miguel
[13]	CARDONA James (+79')
[16]	ORTEGON Wilmer (+81')
[20]	BEDOYA Gerardo
[21]	RESTREPO Oscar

Trainer GARCIA Luis

Cautions
RINCON Freddy (COL) 7', PENA Juan (BOL) 27', BERMUDEZ Jorge (COL) 28', MARTINEZ Gonzalo (COL) 38', DINAS Arley (COL) 55'.

Semi-Structured Data - XML

Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

```
<?xml version="1.0" encoding="utf-8" ?>
<document name="MatchReport.DocID09272.de.xml">
- <MetaData>
    <url>http://fifaworldcup.yahoo.com/02/de/t/q/mr/19326.html</url>
    <language>de</language>
  </MetaData>
- <MatchInfo>
  - <TeamOne>
      <Name>Bolivien</Name>
      <FinalScoreResult>1</FinalScoreResult>
    </TeamOne>
  - <TeamTwo>
      <Name>Kolumbien</Name>
      <FinalScoreResult>1</FinalScoreResult>
    </TeamTwo>
    <FinalResult>1:1</FinalResult>
    <IntermediateResult>1:1</IntermediateResult>
    <MatchDate>26-APR-00</MatchDate>
    <Location>La Paz Hernando Siles</Location>
    <Stadium>12:00</Stadium>
    <Spectator>35500</Spectator>
  </MatchInfo>
- <Goals>
  - <Goal>
      <Scorer>SANCHEZ Erwin</Scorer>
      <Team>BOL</Team>
      <Minute>16</Minute>
    </Goal>
  - <Goal>
      <Scorer>CASTILLO Jairo</Scorer>
      <Team>COL</Team>
      <Minute>32</Minute>
    </Goal>
  </Goals>
- <Cards>
  - <YellowCard>
      <Player>RINCON Freddy</Player>
      <Team>COL</Team>
      <Minute>7</Minute>
    </YellowCard>
  - <YellowCard>
      <Player>PENA Juan</Player>
      <Team>BOL</Team>
      <Minute>27</Minute>
    </YellowCard>
  - <YellowCard>
```

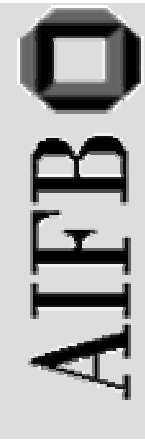
Semi-Structured Data – F-Logic

Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

```
"Bolivien-Kolumbien-26-APR-00_12:00":"http://protege.stanford.edu/kb"#Match.
"Bolivien-Kolumbien-26-APR-00_12:00"
        [
        "http://protege.stanford.edu/kb"#heldOn -> "26-APR-00 12:00";
        "http://protege.stanford.edu/kb"#heldAt -> "La Paz Hernando Siles";
        "http://protege.stanford.edu/kb"#attendance ->35500;
        "http://protege.stanford.edu/kb"#team1 -> "Bolivien-Kolumbien-26-APR-00_12:00_Bolivien";
        "http://protege.stanford.edu/kb"#team2 -> "Bolivien-Kolumbien-26-APR-00_12:00_Kolumbien"
        ].

"Bolivien-Kolumbien-26-APR-00_12:00_Bolivien":"http://protege.stanford.edu/kb"#Squad.
"Bolivien-Kolumbien-26-APR-00_12:00_Bolivien"
        [
        "http://protege.stanford.edu/kb"#players -> Jose_FERNANDEZ;
        "http://protege.stanford.edu/kb"#players -> Juan_PENA;
        "http://protege.stanford.edu/kb"#players -> Marco_SANDY;
        "http://protege.stanford.edu/kb"#players -> Vladimir_SORIA;
        "http://protege.stanford.edu/kb"#players -> Luis_RIBEIRO;
        "http://protege.stanford.edu/kb"#players -> Luis_CRISTALDO;
        "http://protege.stanford.edu/kb"#players -> Victor_ANTELO;
        "http://protege.stanford.edu/kb"#players -> Raul_GUTIERREZ;
        "http://protege.stanford.edu/kb"#players -> Jaime_MORENO;
        "http://protege.stanford.edu/kb"#players -> Ivan_CASTILLO;
        "http://protege.stanford.edu/kb"#players -> Erwin_SANCHEZ;
        "http://protege.stanford.edu/kb"#bench -> Miguel_RIMBA;
        "http://protege.stanford.edu/kb"#bench -> Mauricio_SORIA;
        "http://protege.stanford.edu/kb"#bench -> Joaquin_BOTERO;
        "http://protege.stanford.edu/kb"#bench -> Gonzalo_GALINDO;
        "http://protege.stanford.edu/kb"#bench -> Roger_SUAREZ;
        "http://protege.stanford.edu/kb"#bench -> Miguel_JUSTINIANO;
        "http://protege.stanford.edu/kb"#bench -> Reny_RIBERA
        ].

"Bolivien-Kolumbien-26-APR-00_12:00_Kolumbien":"http://protege.stanford.edu/kb"#Squad.
"Bolivien-Kolumbien-26-APR-00_12:00_Kolumbien"
        [
        "http://protege.stanford.edu/kb"#players -> Oscar_CORDOBA;
        "http://protege.stanford.edu/kb"#players -> Ivan_CORDOBA;
        "http://protege.stanford.edu/kb"#players -> Mario_YEPES;
        "http://protege.stanford.edu/kb"#players -> Alexander_VIVEROS;
        "http://protege.stanford.edu/kb"#players -> Jorge_BERMUDEZ;
        "http://protege.stanford.edu/kb"#players -> Arley_DINAS;
        "http://protege.stanford.edu/kb"#players -> Franki_OVIEDO;
        "http://protege.stanford.edu/kb"#players -> Jairo_CASTILLO;
        "http://protege.stanford.edu/kb"#players -> Gonzalo_MARTINEZ;
        "http://protege.stanford.edu/kb"#players -> Hamilton_RICARD;
        "http://protege.stanford.edu/kb"#players -> Freddy_RINCON;
```



Information Extraction from Free Text

SMARTWEB

Funded by
Federal Ministry
of Education
and Research

Grant 01 IMD01

MATCH-RESULT

FOOTBALL-PLAYER

Linguistic and Semantic Annotation

Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

Mark Crossley saved twice with his legs from Huckerby.

Named Entity Recognition & Semantic Tagging

***[**Mark Crossley **GOALKEEPER] [**saved **GOALKEEPER_ACTION]** twice with his legs from **[**Huckerby **PLAYER]**.*

Linguistic Annotation

***[**Mark Crossley **GOALKEEPER : SUBJ] [**saved **PRED : GOALKEEPER_ACTION]** twice **[**with his legs **PP_OBJ] [**from **[**Huckerby **PLAYER] PP_ADJUNCT]**.*

[GOALKEEPER_ACTION = 'save‘,
GOALKEEPER = 'Mark Crossley‘,
PLAYER = 'Huckerby‘,
MANNER = ‘legs'**]**

Annotation/Extraction Example

SMARTWEB
Funded by Federal Ministry of Education and Research
Grant 01 IMD01

- Example Sentence from Match Report

  Allerdings ist Petrow fuer die Partie gegen Schweden gesperrt und kann erst gegen Ungarn eingesetzt werden.

  “However Petrow has been banned for the match against Sweden and can again be deployed against Hungary.”

- Annotated/Extracted Information

  player_action & [GAME_EVENT "Ban", AGENT player & [SURNAME "PETROW"],
  IN_MATCH game & [TEAM2 "SWE", TOURNAMENT "Match"]]
  team & [NAME "HUN"]

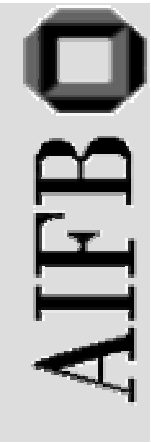
- Use of SProUT (DFKI LT) – Shallow Processing Tool
  - To be Extended with Deep Parsing

Transformation SProUt to F-Logic

SMARTWEB
Funded by Federal Ministry of Education and Research
Grant 01 IMD01

- Transformation of SProUt Output to F-Logic via Declarative Mappings, e.g.:

```
<type orig="player" target="dolce#natual-person-denomination>
  <link type="dolce#natural-person" method="dolce#HAS-DENOMINATION" id=""/>
  <map>
    <simple-mapping>
      <input>
    <arg orig="GIVEN_NAME" target="VAR1"/>
      </input>
      <output method="dolce#FIRSTNAME" value="VAR1"/>
    </simple-mapping>
    <simple-mapping>
      <input>
        <arg orig="SURNAME" target="VAR1"/>
      </input>
      <output method="dolce#LASTNAME" value="VAR1"/>
    </simple-mapping>
  </map>
</type>
```



SProUt to F-Logic

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

```
FS type="player_action">
[N  [N <F name="GAME_EVENT">
      <FS type="world champion"/>
   <F name="ACTION_TIME">
      <FS type="1990"/>
   <F name="ACTION_LOCATION">
      <FS type="Italy"/>
   <F name="AGENT">
      <FS type="player">
          <F name="SURNAME">
             <FS type="Buchwald"/>
          <F name="GIVEN_NAME">
             <FS type="Guido"/>
```

SProUt

```
sobie#player124:sportevent#FootballPlayer
 [sportevent#impersonatedBy ->
   sobie#Guido_BUCHWALD].

sobie#Guido_BUCHWALD:dolce#"natural-
   person"
 [dolce#"HAS-DENOMINATION" ->
   sobie#Guido_BUCHWALD_Denomination].

sobie#Guido_BUCHWALD_Denomination":dolce#
   "natural-person-denomination"
 [dolce#LASTNAME -> "Buchwald";
   dolce#FIRSTNAME -> "Guido"].
```

F-Logic

Display of Extracted Information

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- Automatic Generation of Hyperlink Menus with Extracted Information
  - Ontology-Driven: Hyperlinks will be Added According to Class, e.g. Consider the Hyperlink Menu for the Class **FootballPlayer** (with Instance “Roberto Brown”) :

Das Ziel beider ist klar: Die Qualifikation für die FIFA Fussball-Weltmeisterschaft Deutschland 2006. Diese Aussicht bewegt auch Tejada, der erklärt, dass "es für Panama viel bedeuten würde, zum ersten Mal zu einer Weltmeisterschaft zu fahren. Die Spieler und auch die Menschen im Land würden sich sehr freuen." Was müsste Panama machen, um das Ticket für Deutschland zu holen? Der Goalgetter gibt eine ganz einfache Antwort darauf: "Man muss hart arbeiten. Richtig hart arbeiten ...", wiederholt er.

Der junge Torjäger hofft, auch in den nächsten Partien der abschließenden Sechser-Qualifikationsrunde dabei zu sein, auch wenn er nicht darüber spekulieren möchte. "Das hat der Trainer zu entscheiden", sagt er lapidar. Für die Fans ist Tejada jedoch die Zukunft des panamesischen Fussballs. Bei einer Umfrage auf der offiziellen Website des Fussballverbandes Panamas stimmten 75% für das Sturmduo Tejada und José Luis Garcés, eine weitere junge Sturmhoffnung Panamas. Der altgediente Roberto Brown ist eine weitere Alternative im Sturm.

Roberto Brown
Team: Panama
Events

Goal Event
Costa Rica vs Panama
"Tor von Roberto BROWN (0) in der 58. Minute"
Score: "1:1"

Luis Tejada begann seine Karriere beim FC Tauro de Panamá, bevor er von Envigado (Kolumbien) [illegible] war wohl vorausbestimmt, schließlich hat der Stürmer den gleichen Familiennamen wie ein berühm[illegible] Trainer in der Nationalelf Panamas, Hernández, ist kolumbianischer Herkunft.

Vorbild Ronaldo
Im Fussball der *Cafeteros* ist es ihm bisher gut gegangen. "Ich habe mich verbessern können, natürlich. Es ist eine professionellere Liga (als die Panamas), es herrscht mehr Ordnung, mehr Disziplin, ich habe gelernt, wie ich mich auf dem Platz bewegen muss", erklärt der Spieler. Obwohl es ihm an Erfahrung fehlt, er sehr jung ist und nur wenig Zeit zur Anpassung hatte, war Tejada auf Anhieb in Kolumbien sehr erfolgreich. So hat er in bisher zwölf Spielen dort nicht weniger als sieben Treffer erzielt.

Tejada zeichnet sich als "mannschaftsdienlicher Spieler aus", der "zentral spielt, eine typische Nummer 9 eben, und stark in der Luft ist", obwohl er selbstkritisch meint, dass er sich "im Abschluss noch steigern muss". Als Kind konnte er die Brüder Dely Valdés und Rommel Fernández im Fernsehen bewundern, wenngleich sein Vorbild Ronaldo ist. Mit dem Tor, das Tejada gegen Mexiko erzielte, wäre auch der Brasilianer hoch zufrieden gewesen. Wenn der junge Goalgetter so weiter macht, werden sich die Wege beider vielleicht in Deutschland 2006 kreuzen. Vielleicht wird Luis Tejada dann etwas gesprächiger sein – bislang jedenfalls zieht er ausschließlich die Sprache auf dem Platz vor.

Ontology Learning

Paul Buitelaar, Alexander Schutz (DFKI LT)
&
Michael Sintek, Björn Endres (DFKI KM)
&
Berenike Loos, Robert Porzel (EML)

RelExt - Relation Extraction

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- Motivation: Extend Ontology with Relations
  - Currently ~ 60 Relations in the Sport Events Ontology
    - Mostly Properties, e.g. *hasName, atMinute*, …
  - Representation of (Verbal) Relations Enables Better Modeling of Events for Information Extraction Purposes

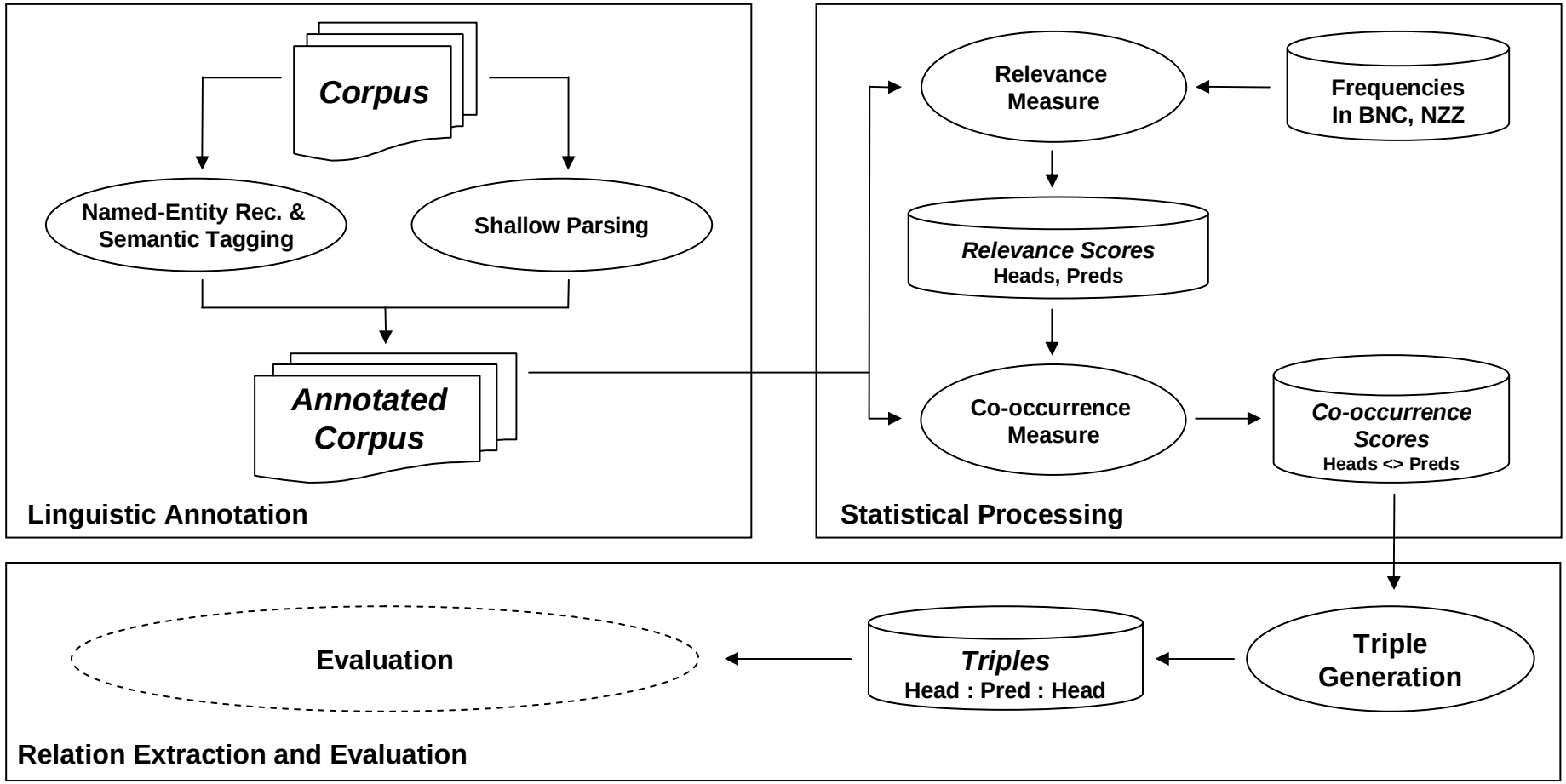
- Example

  "Ballack shoots the ball in the net."

  Relation: Shoot (**Domain:**FootballPlayer **Range:**BallObject)

RelExt – System Architecture

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01



Linguistic Annotation

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- Named-Entity Recognition

  “Michael Ballack” : FootballPlayer

- Semantic Tagging

  “Ball” (ball), “Leder” (leather) : BallObject

- Shallow Parsing
  - Part-of-Speech Tagging

    Fussballspieler (soccer player): Noun

  - Morphological Analysis

    Fussballspieler: Fussball – Spieler

  - Dependency Structure Analysis

    “The team won the second match.”
    SUBJECT PREDICATE DIRECT_OBJECT

Relevance Ranking

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

Rank	χ^2	Headnoun	Frequency
1	125245.24	Ball *(ball)*	6849
2	121888.52	Tor *(goal)*	7767
3	95003.21	Meter *(meters)*	5967
4	64157.18	Schuss *(shot / drive)*	3575
5	57185.76	Ecke *(corner)*	3132
6	45474.96	Strafraum *(penalty area)*	2298
7	34668.11	Freistoss *(freekick)*	1752
8	30017.75	Leder *(leather / ball)*	1561
9	27989.09	Flanke *(cross)*	1479
10	27414.66	Pfosten *(post)*	1457

Rank	χ^2	Concept Label	Frequency
1	565510.99	FOOTBALLPLAYER	28494
2	162137.82	GOALOBJECT	8188
3	143528.88	BALLOBJECT	7249
4	138535.44	GOALKEEPER	6887
5	70814.86	SHOT	3578
6	49018.16	TEAM	2477
7	45474.96	PENALTYAREA	2298
8	34668.11	FREEKICK	1752
9	29324.54	WING	1482
10	28829.78	POST	1457

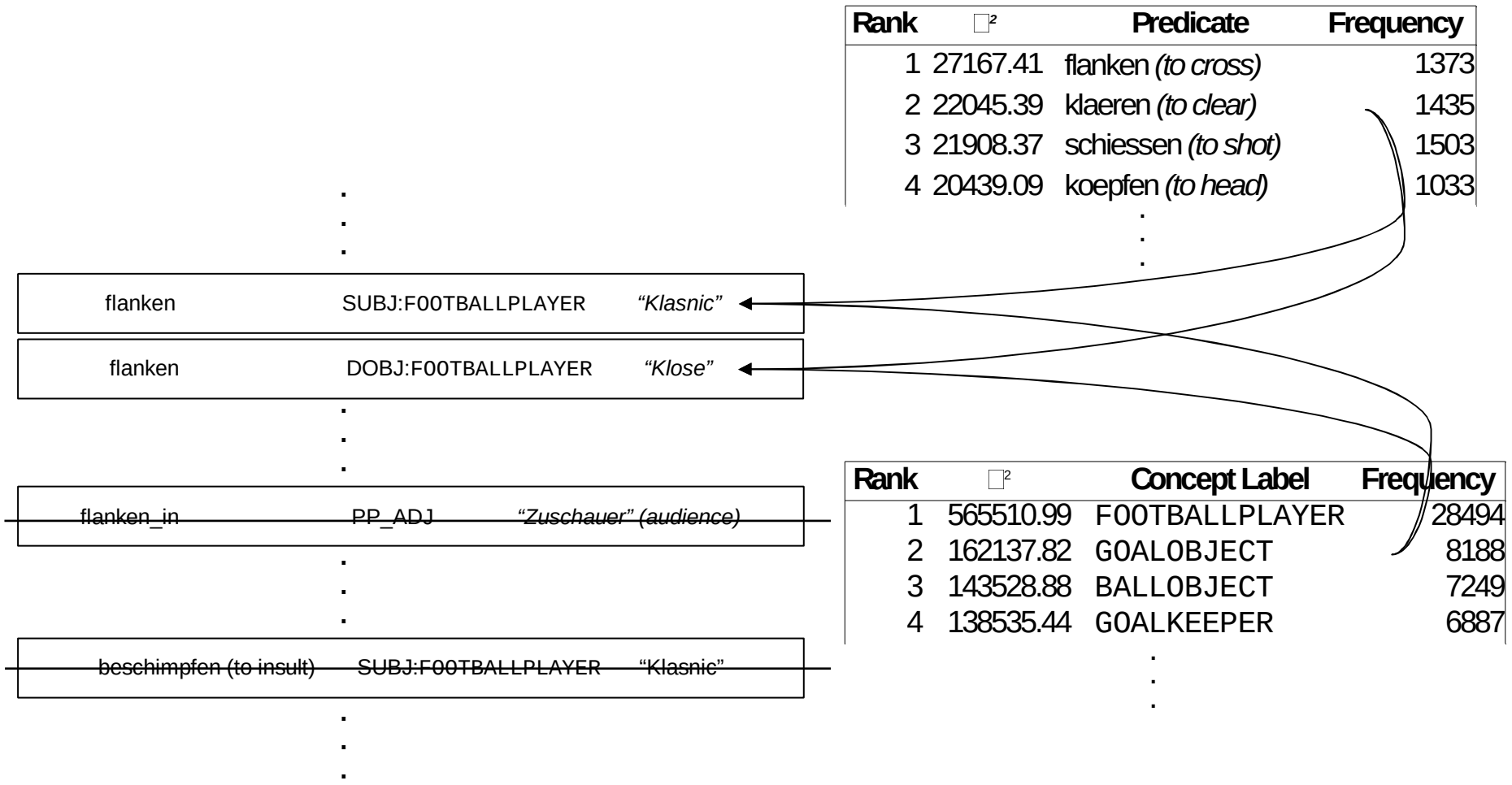
Top-10 Head-Nouns before and after mapping to Ontology Classes

Top-10 Predicates

Rank	χ^2	Predicate	Frequency
1	27167.41	flanken *(to cross)*	1373
2	22045.39	klaeren *(to clear)*	1435
3	21908.37	schiessen *(to shot)*	1503
4	20439.09	koepfen *(to head)*	1033
5	16342.99	lassen *(to let / to leave)*	826
6	9563.41	ziehen *(to pull / to drag)*	1548
7	9468.57	passen *(to pass / to play)*	814
8	7752.84	spielen *(to play / to pass)*	1559
9	7653.68	lenken *(to divert)*	537
10	7637.45	parieren *(to parry / to save)*	405

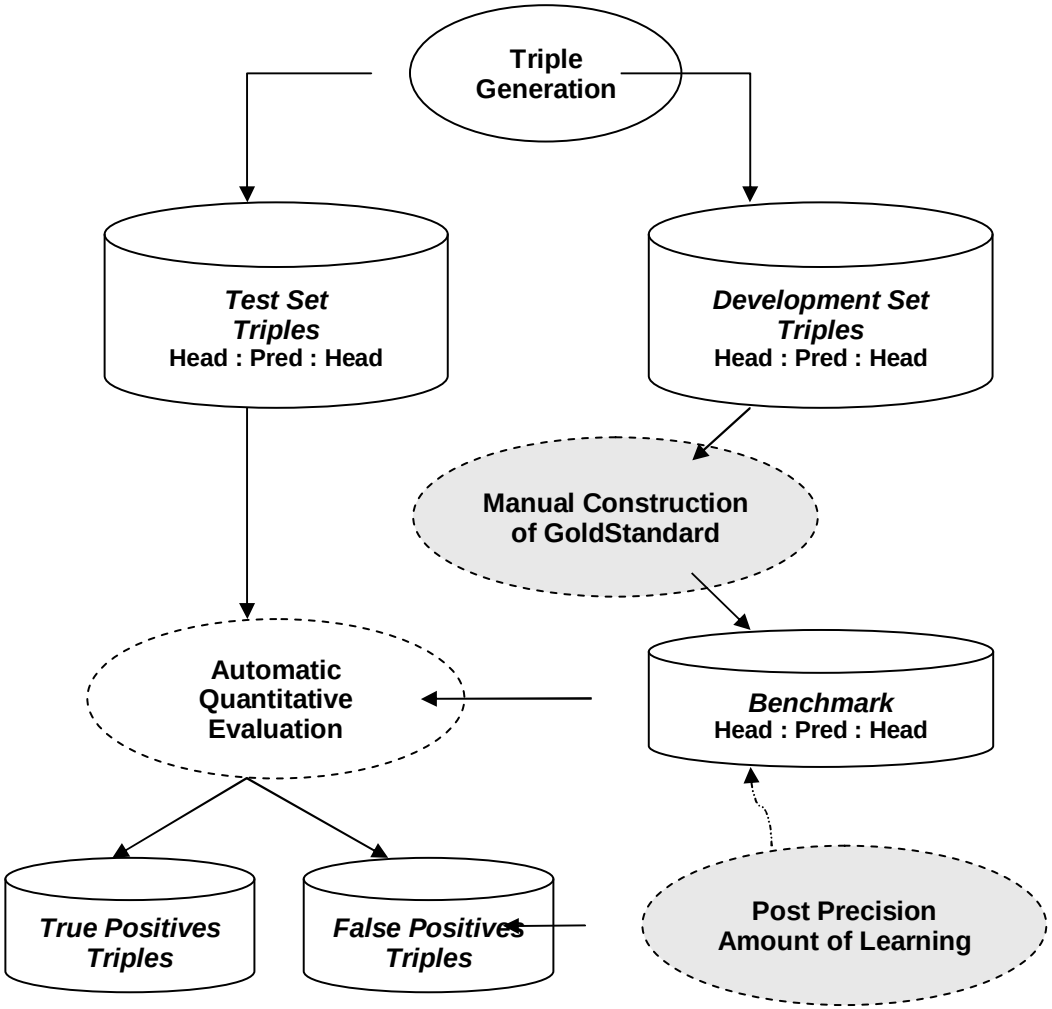
Co-Occurrence Analysis

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01



Experiment – Evaluation

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01



Gold Standard Construction

- 192 Triples from Dev. Set presented to 3 Experts
- Results in a Gold Standard of 38 Triples

Experiment – Results

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

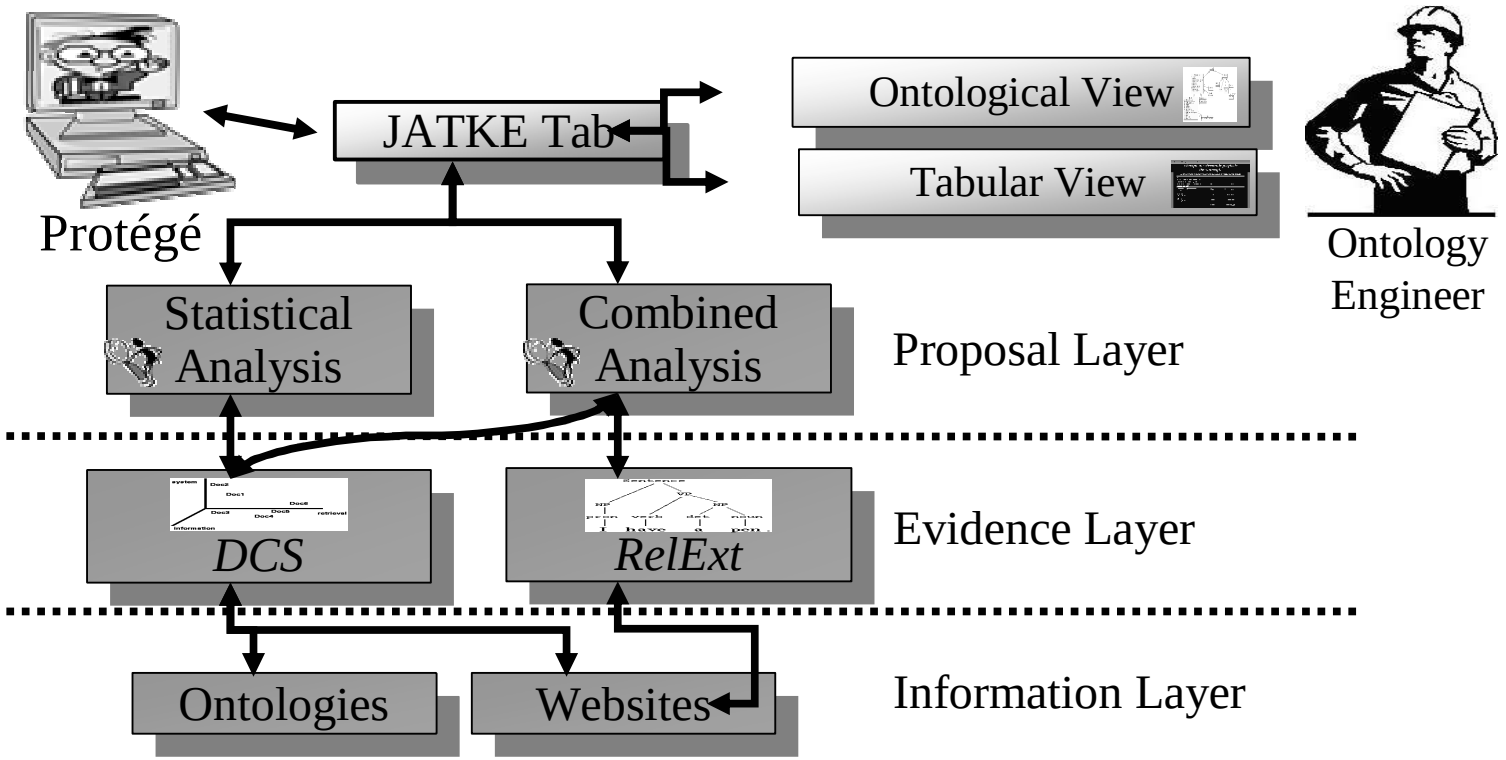
	Corpus	# of Triples Evaluated	Recall	a priori Precision		a posteriori Precision	
				percentage	true positives	percentage	true positives
	1	38	15,80%	8,60%	6	20,00%	14
GS size	2	38	23,70%	13,40%	9	23,90%	16
	3	38	15,80%	8,60%	6	20,00%	14
	Average over Samples		18,43%	10,20%		21,30%	
	1	95	39,50%	12,70%	15	24,60%	29
all Triples	2	84	34,20%	11,90%	13	23,90%	26
	3	92	34,20%	11,10%	13	23,10%	27
	Average over Samples		35,97%	11,90%		23,87%	

	Corpus	# of Triples Evaluated	Recall	a priori Precision		a posteriori Precision	
				percentage	true positives	percentage	true positives
	1	38	13,20%	7,00%	5	18,30%	13
GS size	2	38	21,10%	11,80%	8	19,10%	13
	3	38	15,80%	8,60%	6	15,70%	11
	Average over Samples		16,70%	9,13%		17,70%	
	1	148	44,70%	10,10%	17	20,70%	35
all Triples	2	136	42,10%	10,10%	16	20,30%	32
	3	146	42,10%	9,50%	16	19,60%	33
	Average over Samples		42,97%	9,90%		20,20%	

Integration of *RelExt* in *JATKE*

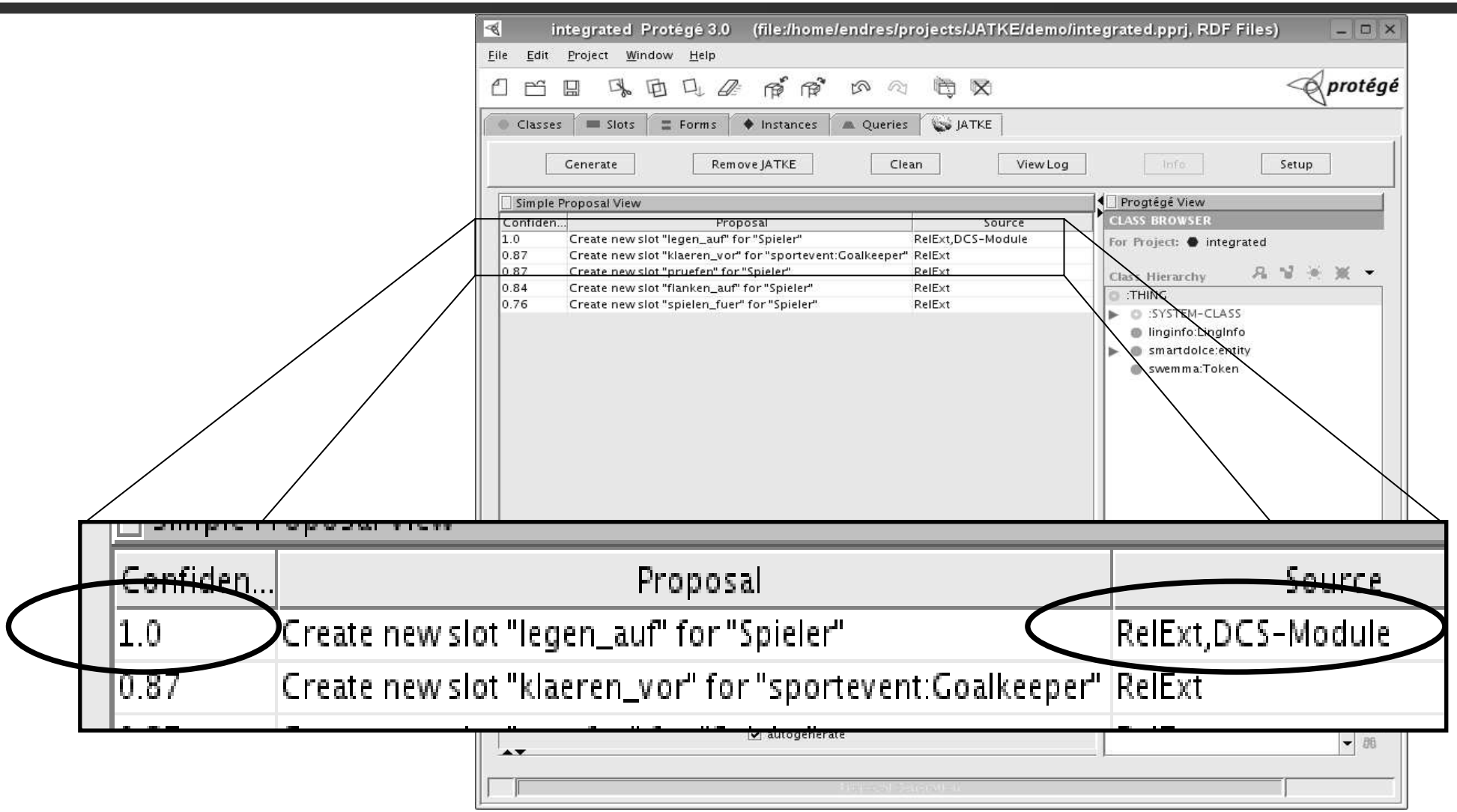
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01

- *JATKE* (Under Development by DFKI KM)
  - Meta-Tool (Protégé Plug-In) for Ontology Learning



Example – Mutual Confirmation

SMARTWEB
Funded by
Federal Ministry of Education and Research
Grant 01 IMD01



Ontology Learning in Dialog (EML) - Motivation

Funded by
Federal Ministry
of Education
and Research
Grant 01 IMD01

How do I get to Auerstein | *Lotus* | *…*

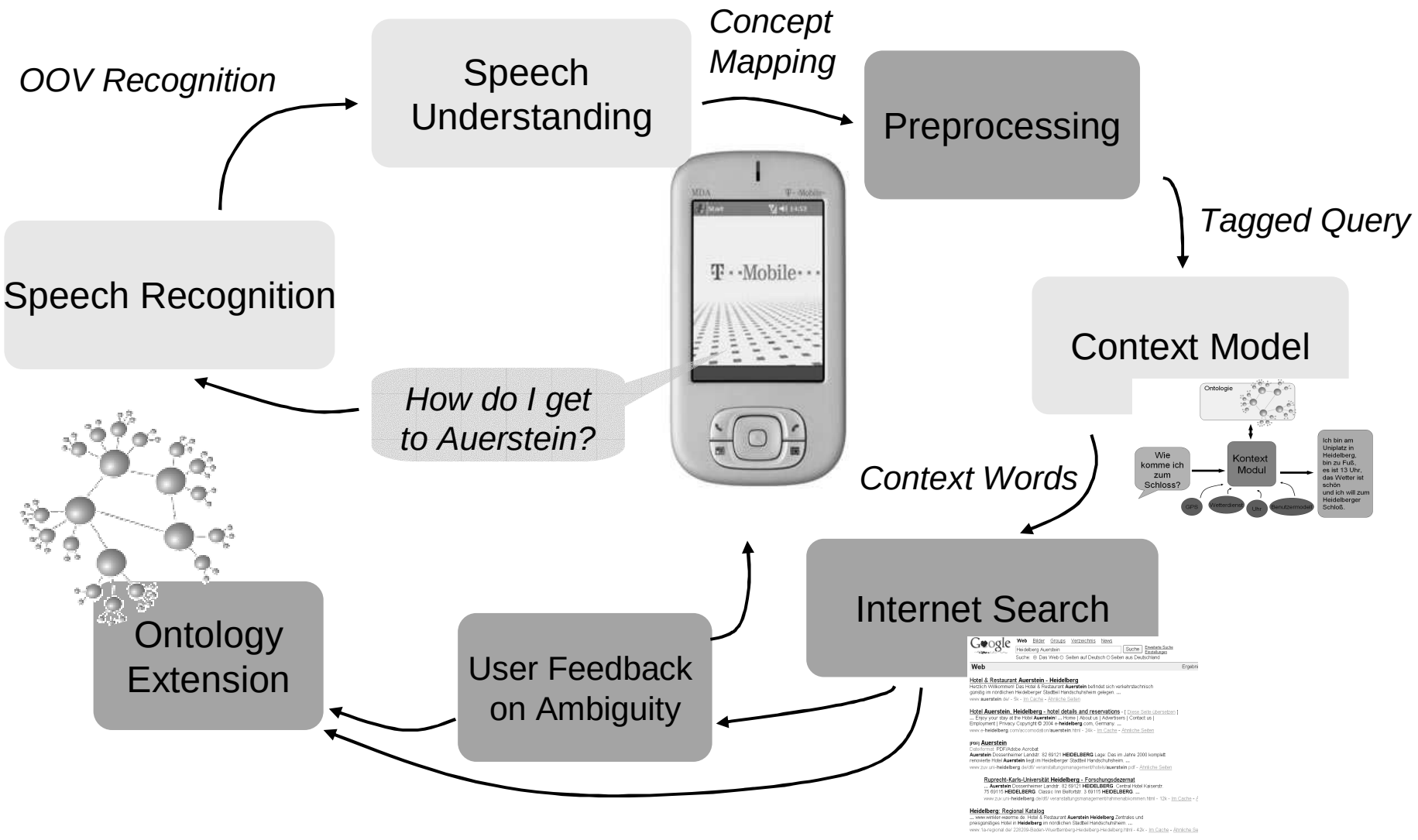
Recognition of ‚Out Of Vocabulary' (OOV) Words:

How do I get to
OOV type=unknown> Auerstein </OOV>
OOV type=unknown> Lotus </OOV>
OOV type=unknown> Betzenberg </OOV>

What was intended? – *Lotus: the car, the flower, the restaurant, …*

Ontology Learning in Dialog (EML) - Overview

Funded by
Federal Ministry
of Education
and Research
Grant 01 IMD01



Funded by
Federal Ministry
of Education
and Research
Grant 01 IMD01

Thanks for Your Attention!