

# An Integration Framework for a Mobile Multimodal Dialogue System Accessing the Semantic Web

Norbert Reithinger, Daniel Sonntag {norbert.reithinger,daniel.sonntag}@dfki.de

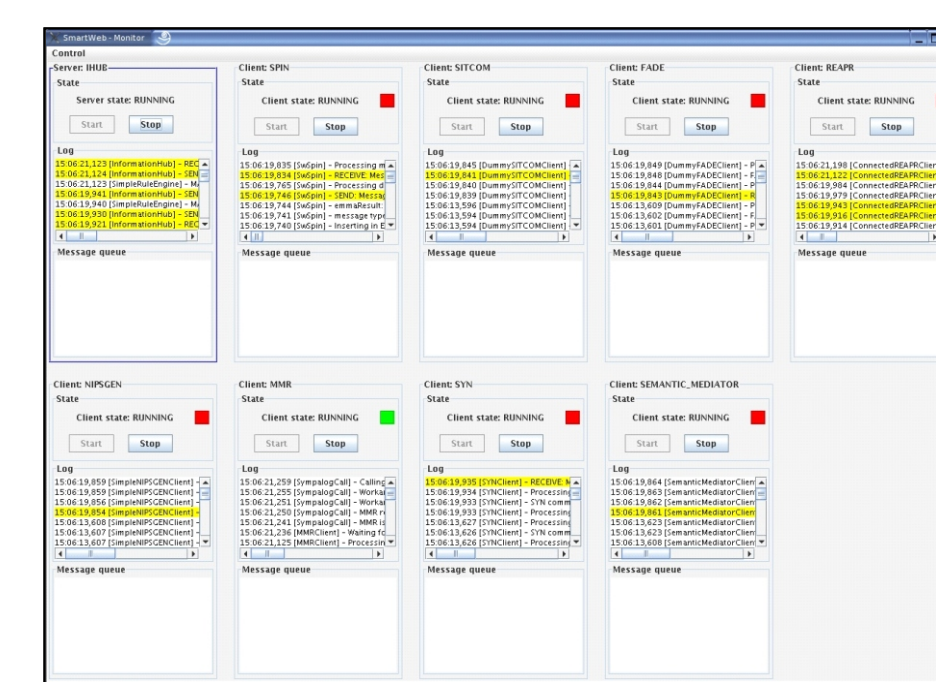
DFKI GmbH, Saarbrücken, Germany

## 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.

## IHUB Architecture

- Simple and robust integration platform for dialogue-internal modules.
- Control of message flow by turn/task version numbers to handle barge-in, reinterpretation.
- IHUB Hub-And-Spoke architecture for easy and fast message transfer.

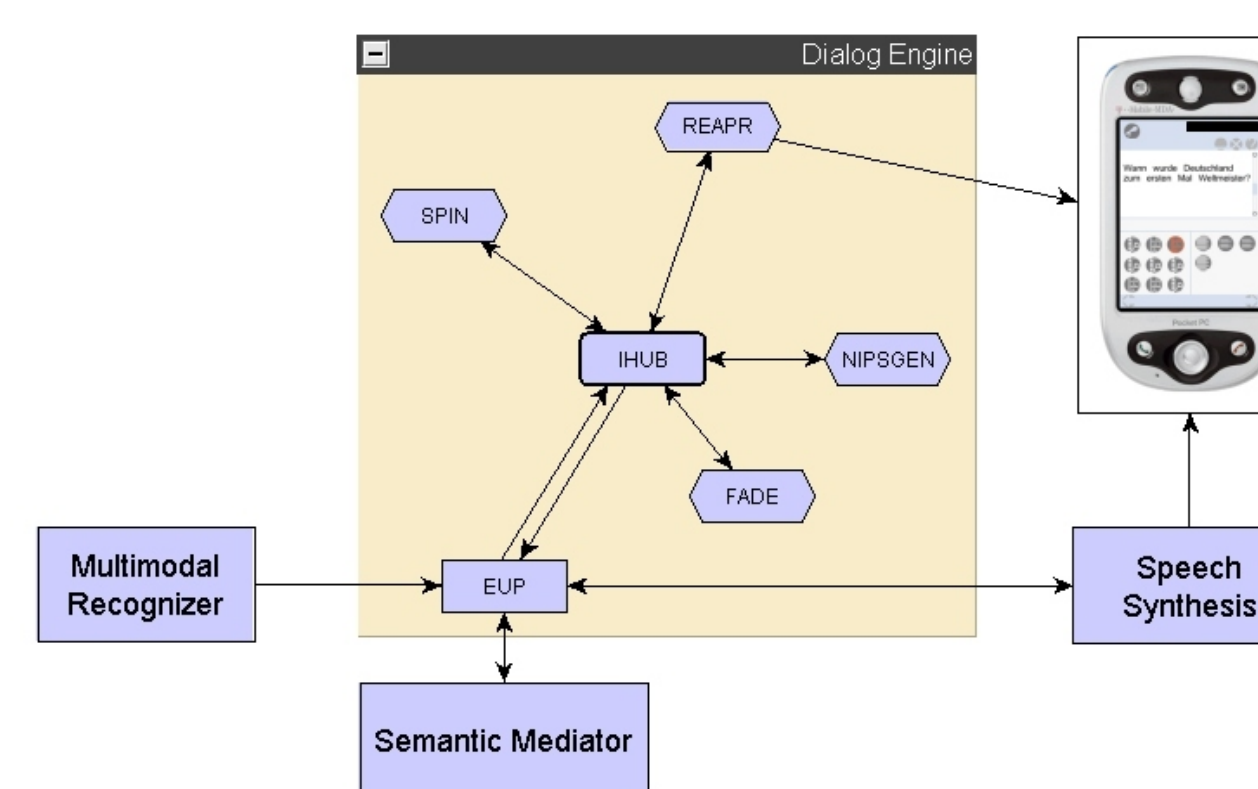


iHUB-GUI

## Components

- Reuse well-tested off-the-shelf components available in Java.
- SPIN: speech interpretation component. Input is mainly the n-best word chains, the output is a parsed word chain.
- FADE: modality fusion and discourse component. Keeps track of the ongoing discourse, completes different types of anaphora and merges input from different modalities. Input is mainly a parsed word chain and the gesture description on an ontological level with geometric position. The output is the enhanced word chain and the completed ontological description of the user's utterance.

Server-based dialogue system with central IHUB



- REAPR: system reaction and presentation module. Decides when a semantic query is accepted for transfer to the Semantic Mediator. Manages interaction, enables user to engage in dialogue processing. Optimises dialogue strategy along with temporal constraints for reaction and result presentation. REAPR is involved in almost all message transfer processes.

## Evaluation

- Formal queries to the Semantic Web are posed in an explicit query language which are ontology instances serialised as EMMA/RDF XML-triples. The user interaction is focussed on deriving this representation from the user input as speech. In combination with the mobile use with its varying acoustic conditions, this requires immediate multimodal feedback on the smartphone.

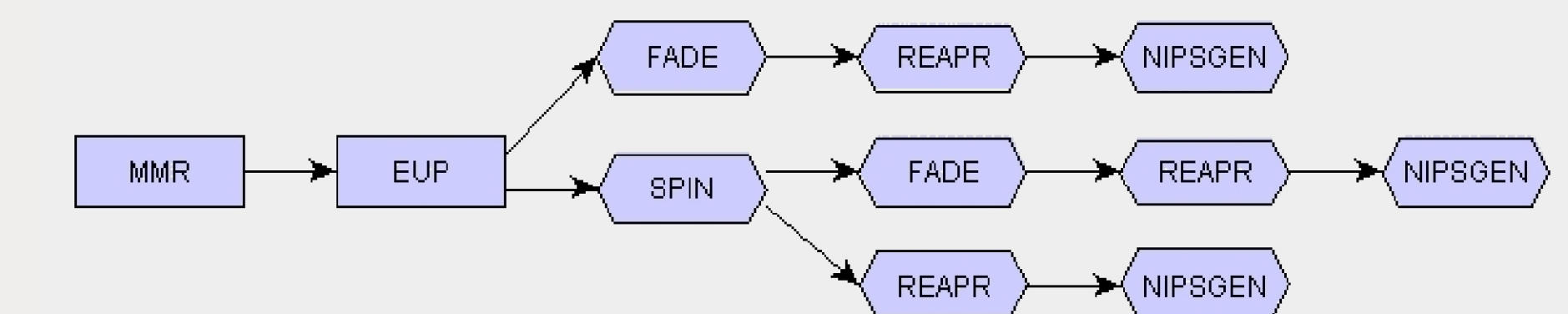
```

Interpretation
confidence = 0.71875
container =
[ Query
  test = "wie sieht der Turnier der Italienischen Nationalmannschaft aus"
  dialogueAct = [ Question ]
  content =
  [ DivisionFootballNationalTeam
    inTournament =
    [ WorldCup HAPPENS-AT =
    [ time-interval BEGINS =
    [ time-point YEAR = "2002" ] ] ]
    people = [ Goalkeeper #02 ]
    origin = [ country IDENTIFIER = "Italy" ]
    focus =
    [ Focus
      focusMediaObject[#02]
      focusObject[#02]
    ]
  ]
]
  
```

- Efficient routing of messages: IHUB automatically discards messages with lower job number, no significant delay compared to module/module communication.

- Actual message content: references to Jena ontological instances.

Message routing test flow for performance evaluation



The exchange of messages containing between 100 and 10000 characters in a realistic message exchange pattern between the components needed between 5.22 ms to 5.72 ms. This time does not vary even when you repeat the experiments up to 10000 times. Medium message run-times between subcomponents over the IHUB for structured data using the Jena data structures are also in the 5ms range on a standard laptop (Pentium 4, 2.66 GHz)

Actual IHUB rules in Smartweb's first demonstrator system

```

(START_CALL,IHUB)-->(REAPR)
(NEW_TURN,REAPR)-->(REAPR)
(START_RECOGNITION,REAPR)-->(MMR)
(RECOGNITION_FINISHED,MMR)-->(REAPR)
(PROCESSING_ERROR,MMR)-->(REAPR)
(JENA,MMR)-->(SPIN,REAPR)
(JENA,SPIN)-->(REAPR)
(JENA,REAPR)-->(SEMANTIC_MEDIATOR)
(JENA,SEMANTIC_MEDIATOR)-->(REAPR)
(NO_RESULT,SEMANTIC_MEDIATOR)-->(REAPR)
(NO_RESULT,SPIN)-->(REAPR)
(NOTHING_SAID,SPIN)-->(REAPR)
(TEXT,REAPR)-->(NIPSGEN)
(TEXT,NIPSGEN)-->(SYN)
(COMMIT,REAPR)-->(*)
  
```

## Special Requirements

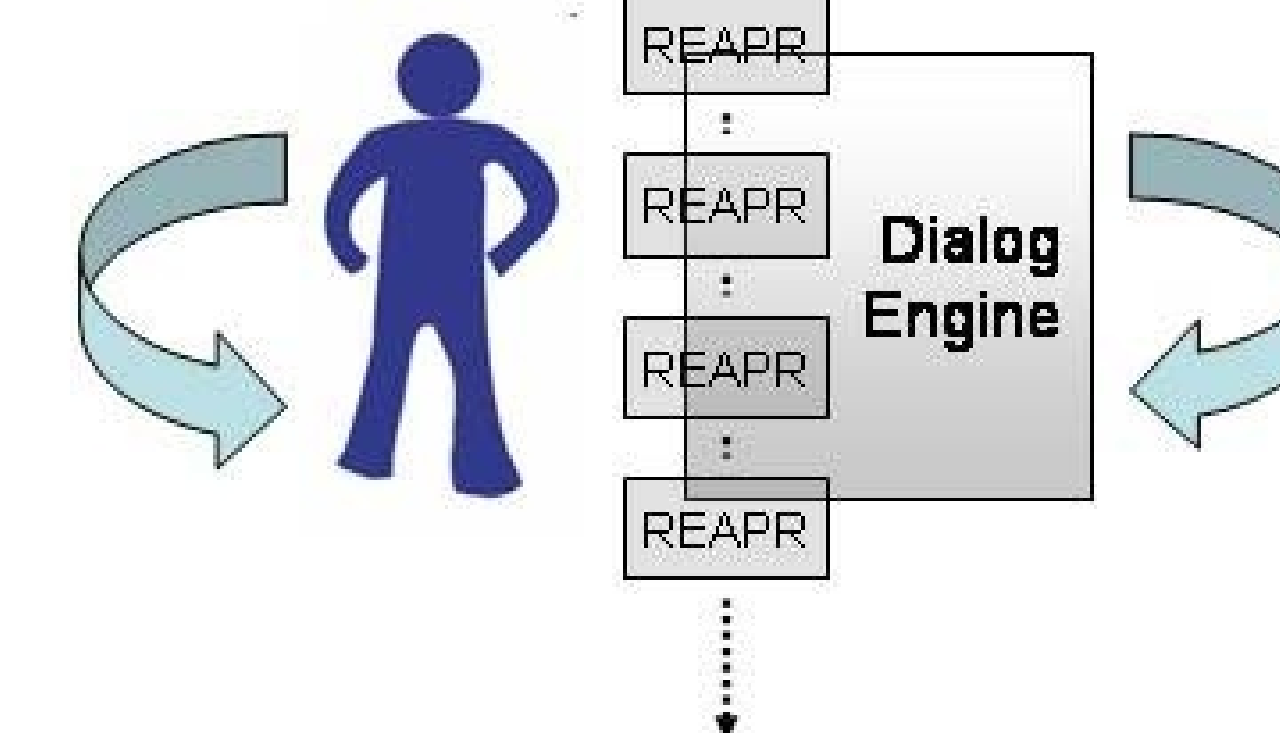
- Develop a context-aware, mobile, multimodal user interface.
- Use a smartphone as interaction device.
- Query transmission via UMTS, WLAN to the backend system.
- Barge-in, multiuser, permanent user control.
- Clear, ontology-based interface between modules (Jena API).
- Real-time interactive editing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

## Standards

- W3C EMMA: for multimodal I/O representation, time stamps, status information, process numbers, generic container for content structures.
- SmartWeb SWEMMA: EMMA Extension for representing queries, results, and status objects.
- RDF for interaction with Semantic Mediator.
- SSML to connect to the Speech Synthesis Subsystem.

## User in the Interpretation Loop

The understood query is displayed for fast editing.



No off-the-shelf solution