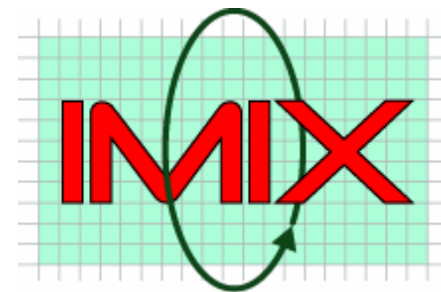


# IMIX

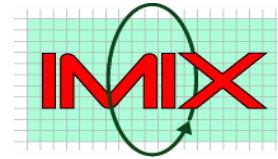
## Interactive Multimodal Information eXtraction

Lou Boves & Gosse Bouma

STEVIN 11 september 2006  
Antwerpen



# Structure



# Programme

***Interactivity +  
Multimodality***

**NORISC**  
automatic speech  
recognition

**IMOGEN**  
multimedia output  
generation

## ***Mediators***

**PARADIME**  
dialogue  
management

**VIDIAM**  
dialogue  
management

**FactMine**  
fact and ontology  
mining

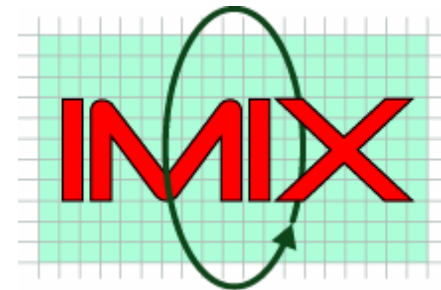
***Information  
eXtraction***

**QA Groningen**  
dependency  
parsing

**QA Rolaquad**  
automatic  
learning

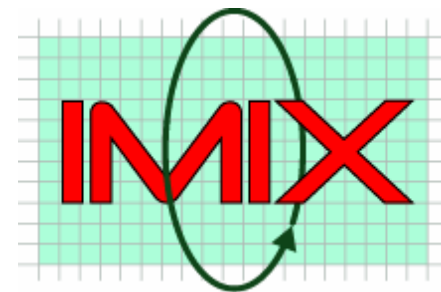
# IMIX focus

- IMIX is een onderzoeksprogramma
  - Ieder project heeft een grote onderzoekscomponent
  - Daarnaast wordt er door alle projecten gewerkt aan een gezamenlijke demonstrator
    - Van kennis, naar modules, naar implementatie in een werkend systeem
    - Interactief vraag-antwoord systeem in het medisch informatie domein



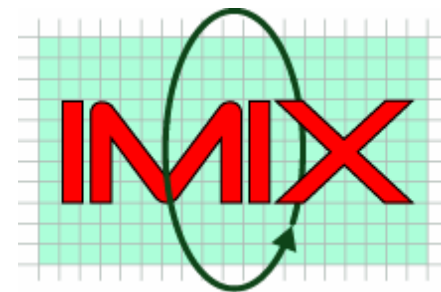
# Uitdaging van IMIX

- Optimalisatie van vraag-antwoordsystemen
- Vraag-antwoordsystemen zo uitbreiden dat er vaker een antwoord gevonden wordt en de antwoorden beter worden
  - Optimalisatie van de input en output op verschillende manieren, o.a. toevoegen van spraakherkenning, plaatjes, filmpjes, talking head
  - Toevoegen van interactiviteit, d.w.z. de mogelijkheid om een dialoog te voeren



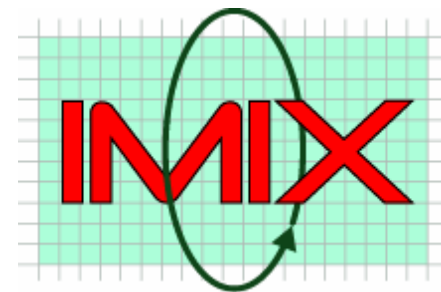
# Vraag-antwoordsystemen

- Optimalisatie door
  - Automatic learning algoritmen (op conceptniveau, bijv. symptomen en definities van een ziekte)
  - Dependency parsing (syntactische informatie gebruiken bij vraaganalyse en antwoord zoeken)
  - Automatisch genereren van ontologie en feitentabellen (waardoor veel antwoorden sneller gegenereerd kunnen worden)
  - Coreferentie-resolutie
  - Intensieve samenwerking tussen de drie vraag-antwoordprojecten



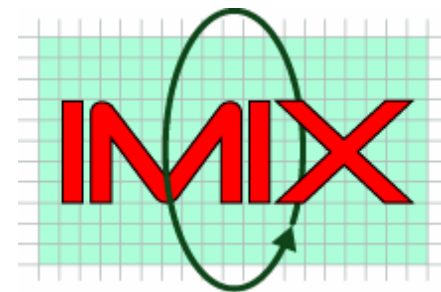
# Optimalisatie van input en output

- Verbeterde akoustische modellen voor de spraakherkenning
- Antwoord-presentatie verbeteren door gebruik te maken van discourse-structuur van door QA gevonden antwoorden
  - Zinsfusering: maakt van 2 antwoorden van 2 vraag-antwoordsystemen 1 beter antwoord
- Toevoeging van plaatjes die ‘bespreekbaar’ zijn: wat is dit? (aanwijzen op plaatje)
- Toevoeging talking head en spraak-uitvoer (Nextens synthese)



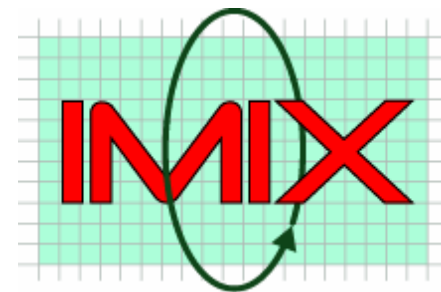
# Interactiviteit

- Toevoegen van een dialoog
  - Clarificatievragen door het systeem
    - Bedoelt u .... of .....
    - Ik kan geen antwoord vinden, zal ik zoeken op....?
  - Follow-up vragen
    - Inclusief aanwijzen d.m.v. pen
- Dit betekent dat er automatisch dialoog acts herkend en gegenereerd moeten kunnen worden



# IMIX en data

- In IMIX zijn een aantal –beperkte-corpora verzameld:
  - Semantisch geannoteerd corpus van medische relaties (oorzaak, gevolg, symptomen, etc.)
  - Kleine (multimodale) dialoogcorpora



# Informatie

- IMIX coördinator, Els den Os
  - O.a. opvragen van special IMIX issue van Dixit ([e.denos@let.ru.nl](mailto:e.denos@let.ru.nl))
- Website
  - O.a. voor gepubliceerde papers  
[www.nwo.nl/imix](http://www.nwo.nl/imix)

