Dialogue management and the visual channel

Projectaanvraag voor een postdoc "Dialogue Management"
NWO IMIX -2de call

Rieks op den Akker
University of Twente
P.O.Box 217, 7500 AE Enschede, The Netherlands
email: infrieks@cs.utwente.nl

February 26, 2004

4 Previous and Future Submissions

THIS IS A FIRST SUBMISSION

5 Institutional Setting

Universiteit Twente,
Afdeling: Informatica van de faculteit Electrotechniek, Wiskunde en Informatica.
Cluster: Taal, Kennis en Interactie (leerstoel: Prof.dr.ir. A. Nijholt).

6 Period of Funding

Startdatum: 1 september 2004
Periode: 1-9-04 tot 1-9-07 (fulltime) of 1-9-04 tot 1-9-08 (parttime)

7 Composition of the Research Team

Hoofdaanvrager	Dr. ir. H.J.A. op den Akker	Universiteit Twente
Postdoc	Dr. ir. B. van Schooten	
Medeaanvragers	Dr. M. Theune	Universiteit Twente
	Dr. J. Zwiers	Universiteit Twente

8 Thematic Priorities

The research fits in the thematic priority 'Dialogue Management in multimodal systems' and will also contribute to the implementation of a multimodal dialogue management system in a distributed blackboard architecture.

9 Description of the Proposed Research Project

9.1 Introduction

The first part of the project concerns the technical specification and implementation of the Dialogue Manager for the (closed domain) medical information-seeking system being developed within the IMIX programme. This part of the project is described in subsection 9.2 The second part of the project concerns innovative research related to one of the challenging aspects of the IMIX system, namely the consequences that the visualization of information provided by the system -in various forms like, pictures, graphics and possibly also movies and animations-, has for the interaction between the user and the system. What is the added value of visual representation, not only for providing the information, but also for querying information, for interaction with the system in general. What are the consequences that follow from these interaction modes for the dialogue manager? This part will be described in subsection 9.3. In section 9.4 we present a meta dialogue system and its role in the incremental development of the dialogue system.

9.2 Aspects of dialogue management in the IMIX system

Dialogue systems can be of various types and we distinguish from simple to more complex types as follows:

QA: A dialogue consists of the user initiating with a question, the system answering the question. (a session can consists of several QA-dialogues; but no information is transfered between dialogues).

QAC: The system can ask for clarification when the initial question is ambiguous -or otherwise unclear- until it can look for information and present the result. There is only one query action in a dialogue. No information is transferred between user-system pairs of actions. (U: "wat is RIS?" S: waar zou u informatie over willen hebben? U: "wat is RSI?" S: "RSI is ...")

DS: After the system has presented information the user can ask a continuation question the system keeps track of the dialogue history; this system will keep track of the topic-structure and should determine topic switch. (U: "wat betekent RSI?" S:"RSI betekent ..." U: "en wat is de oorzaak daarvan?" S:"RSI heeft als oorzaak ...")

DSV: As type DS but extended with the possibility to interact through the visual channel: the user can refer to objects through the visual display - by speech and/or pointing acts and ask for actions to be performed by the system on the display. ("Kunt u dit deel uitvergroten?" "Wat is dat rode bolletje daar?")

The project will develop a dialogue manager incrementally for versions corresponding to the given types QA, QAC, DS and DSV (see "Work Programme"). The tasks of the IMIX demonstrator system as a whole are:

- - to help the user searching the information that he wants.
- - to present the information found in a form that suits best to the users wishes and the type of information.
- - to help the user to identify and query the information that he wants. (presenting the kind of information the system can provide; making suggestions, offering help refering to other information sources)

The data - that is: the information source - consist of a set of annotated documents of a medical encyclopedia, with links to pictures and graphics. The annotation marks the structure of the documents as wel as semantic aspects of the text. The semantic tags refer to key concepts ('disease') and relations between key concepts ('cause') in the medical domain; these are elements of the domain ontology. Synonyms of the key terms are also known.

We first give a very short and simplistic description of the process from user input to output. After that, we discuss the tasks the Dialogue Manager has in this process and the consequences this has for the various modules and for the interaction between the modules. And we sum up the related tasks that have to be carried out in the proposed project.

The user can either use speech or textual input. With spoken input the user can also use deictic acts to point at a picture or to indicate a graphical region. Spoken input by the user is recognized by the speech recognizer that outputs a word graph representing several

possible recognitions accompanied with confidence values. The user query as output by the speech recognizer is tagged semantically and analysed syntactically resulting in a semantical annotated dependency structure or a set of structures (in case of ambiguity). The input module also tries to recognize the speech act (or dialogue act) and in case the input is recognized as a question the system will try to find a set of documents that semantically matches the query. Finally, the user query - the semantical tagged dependency structures - and the documents found (possibly with pictures and graphics) are send to the output module that presents the result to the user, in the form of an answer to the question, by means of written text, speech output or by means of pictures or graphics.

The closed-domain demonstrator the IMIX program aims at is a dialogue system that presents its answers to the user not only in the form of text and speech but also in the form of pictures, graphics or movies.

This 'extension' from a QA-system to a dialogue system introduces some challenging problems related to the key tasks of a dialogue manager.

1. The categorisation and identification of dialogue acts performed by the user.
2. The categorisation and identification of dialogue acts performed by the system.
3. Reference resolution and disambiguation dialogues for reference resolution
4. How do users refer to items that are shown in a picture?
5. Turn taking. Answering a question just by pointing at a picture.

The following decisions with respect to the demonstrators have already been made within the IMIX programme.

- Concerning the user: non-professional user man-in-the-street;
- Concerning the medium: presentation and entrance point for user internet web page
- Concerning the information domain: 'open domain': medical information; closed domain: about RSI.
- Concerning the distributed architecture: multiplatform (see (3))

We will assume that there is a module for knowledge management separate from the dialogue management. The knowledge manager processes information requests and returns answers or error reports in case requests could not be answered. The knowledge consists of the factual information in the documents as well as general knowledge about the closed domain, expressed in a domain ontology. Then there is an input modules that outputs possible interpretations of the input and an output module that presents information or conversational acts performed by the system. Below, we mention the most important aspects of the system design that we consider part of dialogue management or that have direct implications for what dialogue management involves. All these aspects of multimodal interaction can be studied in their own right and research can lead to sophisticated implementations. On the other hand there are also simple solutions that may work quite well; topic change could be signaled by pressing a button for instance. Now, we can formulate the most important jobs as the functional specification of the modules, the integration of the various modules, and the implementation on the the chosen architecture (the Multiplatform (3)).

- communication protocol, turn taking and timing
- multimodal reference resolution
- dialogue act recognition
- system action selection and action planning

Communication protocol and turns The basis of communication is the *protocol* that establishes the common ground for the interaction between user and system. It specifies the rule for correct communicative behavior.
The protocol has to specify a number of rules; they concern:

- when does an interaction start. (that means: when do both system and user agree that an interaction starts)
- when does an interaction stop.
- what are the rules for turn taking/giving

When system or user doesn't behave according to these rules communication is not guaranteed to be successful. This contract is not only the basis that makes communication between user and system possible (at least if the user and the system obey the rules) it is also the basis for the specification of the functionalities that the dialogue systems is expected to have.

The fact that people can talk in their own language to a system makes them expect that the protocol for human human interaction is valid. This may lead to frustrations when this is not the case. And it is never the case; there is always disrupted communication. In the target system there is - to name one disruption - no direct visual contact between user and system, and that excludes the use of mutual gaze as an important communication channel for turn taking. Especially for a system that is for casual use by the layman the interaction protocol must be intuitive; or it has to be made explicit.

Timing is a very important mechanism in interaction, and its effects depend heavily on the modes of communication that can be used. When timing is different from what we expect communication is disturbed, and it is needs some effort to get it back on the rail. Timing is important for synchronisation of different input modalities: synchronicity is an important indication for a semantic association - think about hearing and seeing; pointing and speaking. Timing is important for turn taking: how long does the system wait for a response of the user; it opens the possibility to give semantics to silence (agreement). Does the protocol prescribe orderly turn taking? Or can user or system barge in while the other is speaking? This demands not only parallel input and output processing - something we will presumably have in the IMIX system -, but also incremental input processing, which we will presumably not have. A simple solution for signaling turn giving is when the system (in case of lack of a talking head with eye gaze behavior) uses a visual signal (a flag or red/green light) to indicate when the speaker is 'allowed' to speak. Only field experiments can figure out if this works well.

Multimodal reference resolution Multimodal reference resolution is the process that decides what concepts, objects or actions the user refers to by his referring acts and referring expressions (verbal, deictic) or other means of referring. Reference resolution is one of the central tasks of the dialogue manager. At TKI we developed a system for multimodal interaction (Dutch speech input and pointing on a fixed map by mouse click) in a system for interaction with a navigation agent in a virtual environment (see (4)). For multimodal reference resolution we implemented an augmented version of the Lappin and Leas method using salience of possible referents. The availability of a dependency analyses of user utterances opens the possibility to look at more sophisticated methods for reference resolution than the method that was implemented and in which no semantic roles roles were taken into consideration. A method based on Centering Theory could be considered as well as an extension of a method based on Optimality Theory for multimodal reference resolution by G. Bouma (see (6)). Bouma's experiments were done on the dependency analysis produced by the Alpino parser, that is also used in the IMIX project. Multimodal reference resolution will be an important subject of research when we consider the possibility of the user to interact with the system using the visual channel. We will consider this in more detail below.

Dialogue act recognition The dialogue act recognizer is a module of the dialogue system that has to identify the input of the user in terms of the dialogue act that is intended. In the simple version of the QA system it is assumed that the dialogue act is a question for domain information. The QA module will distinguish several types of this information question. In the extended versions the user can intend a number of different types of dialogue acts. Moreover, when the user uses speech as well as pointing acts in performing a dialogue act, the dialogue act recognizer needs information about the previous acts in order to understand the intention of the user. Hence, dialogue act recognition is a coordinated task between the QA module and the dialogue manager. We assume that we use dialogue act pair information for interpretation of the dialogue act the user intended to perform. Research has shown that knowledge about the last system action gives important indication concerning the expected dialogue act performed by the user (see the PhD thesis of S.Keizer: (7)).

Dialogue acts can be divided in task-related and communication-related. To the last belong various types of feedback. Implicit and explicit verification by the system in case of uncertainty will be a necessary part of interaction with the system. This is boring but necessary because of the expected low performance of general user speech recognition. In the project we propose to do experiments with several versions of the system to determine in an experimental way in what cases implicit or explicit verification and feedback -in which the system reports about its information state- is the best choice.

System action selection On of the main functions of the dialogue manager is to make a decision about the best way to continue a dialogue given the history of the dialogue. Apart from verifications, clarification is expected to be one of the most frequent dialogue patterns that will occur in the information seeking process. It is common to distinguish four different levels of communication (see (1)): signal, semantically understand, intentionally understand, agreement; and on all these levels there can be communication failure (no contact; contact but not enough words recognized; semantically understood but no understanding of the intention in the given context; understood but not agreed), each with their own rules for continuation. The dialogue manager will try to determine as best as possible what type of communication

failure is present in a given situation in order to select the most relevant action to proceed the interaction.

9.3 Visualisation and the Dialogue Manager

In a dialogue the system as well as the user will use the visual channel and both will have particular assumptions with respect to the possibilities the system offers in communication through the visual channel in combination with speech input or output. How does the system use the visual channel? What is the function of a picture? How does the user use the visual channel? How does the system present to the user what possibilities he has to communicate with the system? In what way does the user use expressions that refer to the way something is shown to the user on the graphical display ('Kun je de voet niet wat uitvergroten?' 'Deze voet'; pointing at the display)? This has consequences for other modules of the system as well: the speech recognizer, as well as the context-model. The system should know in what mode and how it presents what kind of information. The system should know what is the visual focus of attention of the user. What are the mechanisms underlying focusing on part of the picture?

Aspects of graphical display that are relevant for referring in dialogue acts are:

- form (bol, rond, plat, dik, ...)
- color (rooie, donkere, ...)
- position (links, achter, boven, ...)
- symbols and icons (pijltje, driehoek)

Request that are related to the visual presentation are for example: "Kun je dat filmpje nog eens afdraaien?" "Op het vorige plaatje zag ik het beter" "Hoe kan ik dit uitvergroten?". The system should know what are the possible modes by which the user can answer a question. Moreover, the user can answer a question by only pointing at some parts, without using speech. Research about the aforementioned forms of use of the visual channel in interaction with the dialogue system will be performed in the project based on experiments with the various versions that will be developed and implemented.

9.4 Methodology

Many decisions that have to be made in developing the dialogue manager - like dialogue strategy (what is the best way to continue the dialogue given the dialogue state; goal selection) and ways of interpretation of user input and presentation of system output - are often based on experiments with similar systems and some further not very well understood intuitions about what seems to be the good choice. However, there is always the question in how far recorded experiences with similar systems can be transferred to the system being developed. We propose to develop the dialogue system by means of an interactive meta dialogue system that presents the informed user (developer of the system) the relevant coordinates of the system state, and offers him a choice from a list of possible continuations (i.e. dialogue acts) that the system sees as possible ways to proceed. The user selects the best one and then takes its turn in response to the systems action it has chosen as the best one. Eventually the system asks the user to evaluate the final state and the dialogue. The system stores the path

followed in the dialogue and stores in every node the state as well as the possible continuations (also those not selected by the user). The user can return to a previous state when he is not satisfied with the dialogue. The choices and selections are stored and could be used to train the dialogue system, learning to select the best strategy, or in determining the design in a next version.
The incremental development of the dialogue system makes it possible to build extended version of the system on evaluation of interactions with previous versions.

9.5 Scientific Relevance

The research in the various aspects of dialogue management in multi-modal dialogue systems for information seeking will contribute to the general knowledge about natural man machine interaction.

9.6 Relevance for the IMIX programme

The dialogue manager is a central component in the multimodal dialog system demonstrators to be developed within the IMIX programme. The interaction via the visual channel has added value in querying information and for the presentation of information in a way that meets the user's wishes.

9.7 Relation with other research activities of HMI (TKI)

The HMI group at Twente University is involved in a number of projects, international, and national, that are related to multimodal interaction and multimodal dialogue systems. Within the European IST projects M4 and AMI we work on: annotation of multimodal interaction, annotation tools and development of annotation schemes for multimodal interaction. The group has experience in development of machine learning techniques for speech act recognition in multimodal interaction.
Web interaction is studied in the PALS project by B. van Dijk and E. Herder and Dutch speech recognition for multimodal information retrieval is one of the research topics of R. Ordelmans. Multimodal presentation in narrative systems and dialogue systems is subject of research by M. Theune, who is also involved in the IMOGEN project within the IMIX programme.

References

[1] H.H. Clark and E.F. Schaefer. Contributing to Discourse. Cognitive Science, 13, 259-294, 1989 Reprinted in: H.B. Clark: *Arenas of Language Use.* The University of Chicago Press, 1992.

[2] A. Flykcht-Eriksson and A. Jonsson. Some empricial findings on dialogue management and domain ontologies in dialogue systems - implications from an evaluation of BirdQuest In: Proceedings of SigDial 2003, Japan, 2003.

[3] G. Herzog, H. Kirchmann, S. Merten, A. Ndiaye, and P. Poller. MULTIPLATFORM testbed: An integration platform for multimodal dialog systems. In H. Cunningham

and J. Patrick, editors, *Proceedings of the HLT-NAACL 2003 Workshop on Software Engineering and Architecture of Language Technology Systems (SEALTS)*, pages 75–82, 2003.

[4] D. Hofs, R. op den Akker and A. Nijholt. A generic architecture and dialogue model for multimodal interaction. In: *Proceedings of the 1st Nordic MUMIN Workshop*, 2003.

[5] M. Nakano, et. al. Handling rich turntaking in spoken dialogue systems. In: *Proceedings of the Sixth European Conference on Speech Communication and Technology* (Eurospeech-99); 1167-1170,1999.

[6] G. Bouma. Doing Dutch Pronouns automatically in Optimality Theory. In: *Proc. of the EACL 2003 Workshop on The Computational Treatment of Anaphora*, Budapest.

[7] S. Keizer. *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*. Twente University Press, 2003.

10 Word Count

3000 words (estimated)

11 International Perspective

This Dutch project contributes to the already existing IMIX-programme. Its international perspective corresponds to that of the IMIX-programme: the development of intelligent multimodal information seeking dialogues to further disclose public available multimodal information.

12 Work Programme

The table shows the project plan over a total period of 36 months (based on fulltime work). The entries in the second and third columns refer to the tasks and to the deliverables described below.

Period	**Task**	**Deliverable**
m1-m4	T1	D1
m5-m12	T2	D2
m13-m20	T3	D3
m21-m36	T4	D4

Tasks

- T1 Implement QA. QA is combined with the visual output generation module (provided by the IMOGEN project) here, though the visual aspect will not be interactive. The main task here is to familiarise oneself with the most important software and provide a first working version of the dialogue system.

- T2 Implement QAC. With the addition of confirmation, we can meaningfully add the speech recognition module at this point. The challenge here is to model a smoothly-functioning confirmation model that includes speech. Evaluation will be necessary, and will be done by means of small traditional experiments measuring recognition rate, time taken, etc.

- T3 Implement DS. Here, we design the main dialogue manager. Evaluation is important at this point. Different alternatives will be evaluated with real users.

- T4 Implement DSV. Here, the main task is extend the dialogue manager to make it interpret visual references in the user's language (which requires an appropriate model of the visual representation) and the user's pointing actions. Evaluation will be done in a similarly extensive manner as T3.

Deliverables

- D1 The first version of the demonstrator

- D1 The second version of the demonstrator

- D1 The third version of the demonstrator

- D1 The fourth and final version of the demonstrator

Specificeer en faseer het werkplan voor de beoogde subsidieduur. U wordt verzocht de workpackages, de beoogde deliverables (artikelen en/of software), de tijdplanning en de onderzoekers die de deliverables moeten opleveren samen te vatten in een overzichtelijk schema of in een tabel. U dient gedurende de gehele subsidieperiode tijd te reserveren voor het schrijven van publicaties. In het werkprogramma dient u ook voldoende tijd te reserveren voor de integratie van de resultaten van het onderzoek in (n van) de IMIX demonstrator(s). U wordt verzocht expliciet aan te geven op welke momenten in de tijd welke resultaten voor integratie beschikbaar komen, en welke additionele werkzaamheden nodig zijn voor een daadwerkelijke integratie.

13 Planned Deliverables and Knowledge Dissemination

For deliverables see the table in the previous subsection. Research results will be presented at the national and international conferences and workshops on multimodal interaction and dialogue systems. Dissemination will mostly be a joint activity with the other partners in the IMIX programme.

14 Short CV Principal Applicant and Candidate Postdoc

Principal Applicant

Title(s), name:	Dr. ir. Henderikus J.A. op den Akker
Birth:	10 january 1952 te Leeuwarden
Nationality:	Dutch

Rieks op den Akker studied mathematics, and computer science and did his Ph.D in computer science (title of thesis: "Parsing Attribute Grammars", 1989) at the University of Twente, Enschede, the Netherlands, where he is now assistant (UD) of Prof. dr ir. A. Nijholt. His research interest are: multimodal interaction and natural language dialogue systems for interaction with virtual environments. He contributed to several publications in this area. He is involved in European projects (M4 and AMI) and in the national project Pidgin (Senter-Min.van Econ.Zkn. "speerpuntonderzoek").

Candidate Postdoc

Title(s), name:	Dr. Ir. Boris Wessel van Schooten
Birth:	7 august 1972 te Almelo
Nationality:	Dutch

Boris van Schooten studied computer science at the university of Twente, and got his master's degree in april 1998. Following this he did his Ph.D. at the same faculty, within the TKI research group. His supervisor was Anton Nijholt. He obtained his Ph.D. degree on 17 april 2003. The title of the thesis is "Development and specification of virtual environments". The Ph.D. research identifies four main characteristics of virtual environments (VEs): graphical interaction, multimodality, interface agents, and multi-user. It proposes a general systems development framework regarding these characteristics, and proposes specification languages and a software toolkit for building VEs.

15 Literature

Selection of publications

References

[1] R. op den Akker and A. Nijholt. Dialogues with embodied agents in virtual environments. In: *Proceedings NLP 2000*, Lecture Notes in AI, Springer, Berlin, 2002

[2] R. op den Akker, M. Hospers, D. Lie, E. Kroezen and A. Nijholt. A rule-based reference resolution method for Dutch discourse. In: *Proceedings 2002 Symposium on Reference Resolution in Natural Language Processing*, University of Alicante, Spain, S. Harabagiu and A. Ferrandez (eds.), June, 2002, 59-66.

[3] B. van Dijk, R. op den Akker, J. Zwiers and A. Nijholt. Navigation assistance in virtual worlds. In: Informing Science, Special Series on Community Informatics, E. Rathswohl and C. Winer (eds.), ISSN 1521-4672, 2003, Vol. 6, 2003, 115-125.

[4] D. Hofs, R. op den Akker and A. Nijholt. A generic architecture and dialogue model for multimodal interaction. In: *Proceedings of the 1st Nordic MUMIN Workshop*, 2003.

[5] Hospers, M. and Kroezen, E. and Nijholt, A. and Op den Akker, R. and Heylen, D. An agent-based intelligent tutoring system for nursery education. In: *Applications of software agent technology in the health care domain* (Antonio Moreno, John L. Nealon, eds). Birkhauser Verlag, Basel, 2003

[6] S. Keizer, R. op den Akker and A. Nijholt. Dialogue Act Recognition with Bayesian Networks for Dutch Dialogues. In: *Proceedings 3rd SIGdial Workshop on Discourse and Dialogue*. K. Jokinen and S. McRoy (eds.), Philadelphia, Pennsylvania, July, 2002, 88-94.

[7] M. Poel, R. op den Akker, A. Nijholt and A.J. van Kesteren. Learning emotions in virtual environments. In: *Cybernetics and Systems 2002*. R. Trappl (ed.), Austrian Society for Cybernetic Studies, ISBN 3-85206-160-1,Vienna, 2002, 751-756.

International literature

References

[1] A. C. Graesser, N. Person, and D. Harter. Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, 2001. in press.

[2] G. Herzog, H. Kirchmann, S. Merten, A. Ndiaye, and P. Poller. MULTIPLATFORM testbed: An integration platform for multimodal dialog systems. In H. Cunningham and J. Patrick, editors, *Proceedings of the HLT-NAACL 2003 Workshop on Software Engineering and Architecture of Language Technology Systems (SEALTS)*, pages 75–82, 2003.

[3] M. Johnston, P. R. Cohen, D. McGee, S. L. Oviatt, J. A. Pittman, and I. Smith. Unification-based multimodal integration. In *ACL '97/EACL '97: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and of the 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 281–288, 1997.

[4] Arne Jönsson and Magnus Merkel. *Some issues in dialogue-based question-answering.* In *Working Notes from AAAI Spring Symposium*, 2003.

[5] Jean-Claude Martin. Towards 'intelligent' cooperation between modalities: the example of multimodal interaction with a map. In *Proceedings of the IJCAI'97 workshop on Intelligent Multimodal Systems*, 1997.

[6] Christof Monz. From document retrieval to question answering. PhD Thesis; ILLC Dissertation Series DS-2003-4, University of Amsterdam, Amsterdam, 2003.

[7] Wolfgang Wahlster, Norbert Reithinger, and Anselm Blocher. Smartkom: Towards multimodal dialogues with anthropomorphic interface agents. In *MTI Status Conference*, 2001.

[8] R. Catizone, A. Setzer, Yorick Wilks. Multimodal dialogue management in the COMIC project. In: Proc. of the Workshop on Dialogue Systems. EACL 2003, pp.25-34.

[9] A. Kerminen and K. Jokinen. Distributed dialogue managment in a blackboard architecture. In: Proc. of the Workshop on Dialogue Systems. EACL 2003, pp.53-60.

[10] P. Lendvai, A. van den Bosch and E. Krahmer. Machine learning for shallow interpretation of user utterances in spoken dialogue systems. In: Proc. of the Workshop on Dialogue Systems. EACL 2003, pp.69-78.

16 Samenvatting

In het IMIX programma wordt onderzoek gedaan naar verbeteringen van technieken voor vraag-antwoord systemen waarbij het Nederlands zowel de vraagtaal is als de taal van de documenten waarin de informatie wordt gezocht. Het gaat daarbij om een aantal aspecten: de interpretatie van de invoer, waarbij ook gesproken invoer wordt onderzocht, het zoeken van de mogelijk relevante stukken tekst, het formuleren en presenteren van een antwoord op basis van de gevonden teksten en van eventuele plaatjes of grafieken.
Met een vraag-antwoord systeem is een eenvoudige dialoog mogelijk: de gebruiker stelt een vraag; het systeem geeft een antwoord. Maar stel nu dat het systeem de vraag niet begrijpt. Hoe moet het dan reageren? Moet het suggesties geven? Duidelijk maken waarom het iets niet begrijpt? Het zou handig zijn wanneer in een dialoog tussen gebruiker en systeem de gebruiker geholpen kon worden bij het zoeken van informatie en het formuleren van de vraag zodat het systeem aan de informatiebehoefte kan voldoen. In dit onderzoekproject gaat het om een bijdrage te leveren aan een dialoog-systeem voor het zoeken van informatie in min of meer ongestructeerde teksten in een beperkt medisch domain, waarbij de gebruiker zowel spraak, tekst als aanwijzen als middelen kan gebruiken om z'n informatiewens kenbaar te maken en waarbij het systeem de antwoorden zowel in tekst, spraak als beeld kan presenteren. De visualisatie van informatie in de vorm van plaatjes, grafieken en animaties opent nieuwe mogelijkheden voor interactie, en nieuwe manieren om informatie op te vragen. De gebruiker zal daarbij refereren naar zichtbare objecten op een wijze die te maken hebben met de grafisch vormgeving zelf. Dit stelt bijzondere eisen aan het systeem: het moet weten op welke wijze het objecten of informatie grafisch voor de gebruiker presenteert, waar de aandacht van de gebruiker tijdens de interactie op gericht is en hoe deze refereert naar objecten en informatie onder invloed van de wijze van presentatie. Dit heeft gevolgen voor alle aspecten van dialoogmanagement. Die gevolgen worden in kaart gebracht en er wordt gezocht naar praktische methoden die interaktie via het visuele kanaal in een multimodaal dialoog systeem mogelijk maakt.

17 Research Budget

Het budget dat wordt aangevraagd voor dit project is conform the standaardafspraken die gelden volgens het NWO-VNSU-reglement.
Daarbij zij aangetekend dat de beoogde Postdoc te kennen heeft gegeven voorkeur te hebben voor een parttime aanstelling (75 percent) gedurende 4 jaren.

HMI is involved in three BSIK projects: MultimediaN, ICIS, and BiOrange. The contribution of HMI to these projects concerns, among others, multimodal interaction and design of (adaptive) user interfaces. Therefore the research proposed here will (also) be performed in the framework of the BSIK programme (formulation suggested by NWO).