Formative Evaluation of the SmartWeb Prototypes

Version 1.8

Hannes Mögele, Florian Schiel

Ludwig-Maximilians-Universität Munich

Technical Document No. 11
September 2007

Formative Evaluation of the SmartWeb Prototypes

September 2007

Hannes Mögele, Florian Schiel

IPS LMU München
Schellingstr. 3
80799 München

Tel.: (089) 2180-5751

E-Mail: schiel@bas.uni-muenchen.de;
hannes@bas-uni-muenchen.de

This technical document belongs to sub-project 9.1:
Evaluierung

The technical document belongs to a research project that was supported with funding from the Federal Ministry of Education and Research under the funding number 01 IMD 01. The responsibility for the content lies with the authors.

Table of Contents

1 Introduction

This document gives a brief description of the formative evaluation that was performed on the regular releases of the SmartWeb prototypes starting with version 0.5.1 (Oct 2006) until the pre-final version 0.9 (June 2007).
The SmartWeb system is a server-based, multimodal, spoken dialogue system for accessing Internet-based resources with multimedia features. The client software runs on a mobile handheld device MDA Pro under Windows Mobile. The speech recognition, speech synthesis, query extraction, content extraction and presentation design is handled by a server application. The spoken and textual access is domain-independent; the recognition dictionary is in principle unrestricted. Connectivity between server and client is achieved by standard wireless Internet protocols such as WLAN and UMTS. Although the handheld device, the software application on client and server and the service providers form an interdependent unit the object of this evaluation is only the SmartWeb application (server and client) itself.

The technique of this evaluation is roughly based on the decisions of the Evaluation-Workshop in Munich (22. June 2006) - see the appropriate protocol - and on correspondence with interested partners, mainly DTAG. In this TechDoc we do not document the decision process for the evaluation but merely describe how the formative evaluation of the developed SmartWeb prototypes 0.5.1, 0.6, 0.7, 0.8 and 0.9 were realized and summarize the results which are presented in the appendix to this document in full.

The test methodology was roughly as follows:
10 test subjects were chosen to run a series of tests up to a maximum of 10 task levels. Each level consisted of a pre-defined task that the test subject was asked to solve with the aid of SmartWeb (see section 2.2). The first test session (level 0) was designed as a training session where test subjects learned how to use SmartWeb and the MDA Pro by uttering pre-fabricated queries from a list.[1]

Each new SmartWeb prototype release was installed immediately on our server platform in Munich (see section 2.1) and tested for technical functionality. Then a standard test was performed by an expert to verify that a required minimum of SmartWeb functions were available in the new release. After passing this pre-test the already running formative evaluation tests were switched to the new released prototype.
For example, a test subject might have performed the first 3 test sessions with prototype 0.5.1 and then switched to prototype 0.6 with the fourth test session. Since the release dates were unknown, the distribution of test session levels to prototypes is not controlled in this evaluation. However, it is possible to track evaluation results along a test session series of a single test subject and measure the improvement of the performance.

All test conditions including meta data of test subjects (see section 2.3) as well as log files of the system and results of questionnaires were gathered in a common evaluation database (see section 2.4).

The motivation for not using one test subject per test session is based on our experiences with the SmartKom evaluation performed in the years 2003 and 2004. Naive test subjects need time to adapt to a new technology; it is not realistic to evaluate the acceptance and ergonomics of a new technology within a period of 1 hour. Also, we were interested in the

[1]Also, this training session enabled us to verify that the chosen test subject was no 'goat' (that is, a person where for unknown reasons the speech recognition yields only very bad results). No 'goats' were encountered during the formative evaluation.

general attitude of test subjects towards dialogue systems after being exposed to the SmartWeb prototypes as well as their judgment about improving technology (with each new prototype release), that is will the test subjects notice any improvements after an prototype update.

2 Basic Test Setup

In this section we briefly describe the overall design of the evaluation experiments including the technical setup, the logistics of a test series, the controlled test conditions as well as the recorded evaluation data.

2.1 Prototype Technique

Server

The server application of all SW prototype releases was installed on two Intel based hosts. The technical data of these identical hosts are:

- P4 2,80GHz
- 3GB RAM
- 100 Mbit LAN
- OS: Linux SuSE 9.2 (0.5.1) SuSE 10.1 (> 0.6)

On the first host 'host1' the SW testbed is started with the option 'services':

```
testbed.sh services
```

and after 15 sec the SW testbed is started on the second host 'host2' with option 'dialog':

```
testbed.sh dialog host1
```

'host1' is therefore the server address used by the clients.

Connectivity to the clients is realized either per UMTS (German T-Mobile) or via the in-house WLAN network using 802.1X protocol; for this purpose a driver of secureW2 was installed on all used MDA Pro clients.

Client MDA Pro

A total of five identical MDA Pro were used for the evaluation tests:

Manufacturer :	T-Mobile
Mobile type :	Twist-and-Flip
Category :	Smartphon
Networks :	GSM900/GSM1800/GSM1900/W-CDMA(UMTS)
Operating System :	Microsoft Windows Mobile 5.0
ROM Version :	ROM 1.30.113

The SmartWeb client software as well as required packages were installed using ActiveSync. No other software was allowed on the clients than the following packages:

- IBM PPRO10
- IBM PPRO10 (German)

- Microsoft .NET CF 2.0
- Macromedia Flash Player ActiveX
- SecureW2
- SmartWeb

After installation the following permanent settings were set on each client:

- AGC off
- Energy savings display: permanently on for battery, 5 min off for charging
- Energy savings system: permanently on for battery, 5 min off for charging

In the SW ConfigEditor the following permanent settings were set on each client:

- A/V: all of (except for special sessions with on/off-focus tests)
- language : de
- GPS : either Berlin or Munich (depending on the task the appropriate unit was selected)
- profile / default : settings not changed

Most experiments were conducted with the delivered standard T-Mobile-Headset for the MDA Pro that was cable-connected; there was no use of the Bluetooth connection. The tests were either performed indoors (office, cafe) or outdoors (cafe, open street). The outdoor tests took place under different conditions of weather and noise levels. For experiments with on/off-view enabled the MDA Pro was used with the main camera directed to the users face (open position).

Tested Releases

The following table shows the tested SW prototype releases together with their subversion number, date of installation, pre-test result and applied operation system (OS)

Release	Subversion	Date	Pre-Test	Server OS	# of tests
0.5.1	6520	2006/11/23	ok	SuSE 9.2	17
0.6	7214	2007/01/09	ok	SuSE 10.1	52
0.7	7585	2007/02/19	failed	SuSE 10.1	-
0.8	7780	2007/04/03	ok	SuSE 10.1	21
0.9	8118	2007/05/11	failed	SuSE 10.1	-

Note that releases with failed pre-test were not included in the formative evaluation. Two test were performed by a test subject that was not able to complete the whole series and were therefore excluded from the later analysis.

2.2 Task Concept

The evaluation tests were structured into max. 10 task levels starting with a training task on level 0. Task levels are synchronized and ordered, that is, no test subject performed a task

of level 3 before having performed the pervious levels 0,1 and 2. Due to the reduced project time not all test subjects completed all 10 task levels.

For each task level the test subject received a verbal instruction by the investigator as well a printed document describing the task in detail. The tasks and their general topics are:

Task Level	Topics	Environment	Number of tests
0	Training: Usage of MDA Pro, Listed Queries	office	10
1	Watching Soccer with Friends, World Series 2006	office / cafe	9
2	Planing a short visit to a German City	office / cafe	10
3	Planing a Visit to the Cinema I	office / cafe	10
4	Visit to Berlin	office	10
5	Planing a Visit to the Cinema II	office / street	10
6	List of pre-defined queries	office / street	10
7	(Free selection of Topic as in Level 1-4)	office / street	8
8	Looking for a good physician	office / street	6
9	Free queries, personal interests, typical search queries as given to Internet search engines	street	5

A print-out of the original task tests (German) can be found in appendix 8.5.
Task level 3 and 5 were chosen to be identical to search for test subjects adaptation effects. No such adaptation effects were detected. Therefore we can assume that all test subjects adapted very rapidly to the new technology within the first 2 task levels.

2.3 Test Protocol

The following table shows the most prominent test conditions and their possible values for all experiments.

Condition	Possible Values	DB Key
Session Number	e[0-9][0-9][0-9]	s_session_name
Test subject code	[A-Z][A-Z][A-Z][A-Z]	s_spkcode
Date	YYYY-MM-DD	s_recordingdate
Investigator	<free text>	s_investigator
Start time	HH:MM:SS	s_start
End time	HH:MM:SS	s_end
Client	MDA_Pro_[A-E]	s_mobiletyp
Headset	MDA_Pro, none	s_headsettyp
Push-to-talk	ptt, active	s_asr_state

Condition	Possible Values	DB Key
Environment	indoor, outdoor	s_situation
Subject is walking	yes, no	s_walk
Location	office,street,cafe	s_directions-location
Background noise	yes, no	s_background_noise
Weather	sunny, cloudy, rainy	s_weather
Age of subject	<number>	v_age
Props	<free text, e.g.'backpack', 'handbag',>	v_props
Visible piercings	yes, no	v_piercing
Subject wears Glasses	yes, no	v_glasses
Subject is bald	yes, no	v_bald_head
Subject has a beard	yes, no	v_beard
Smoker	yes, no	v_smoker
Experience with dialogue systems	yes, no	v_dialog_system_experience
Experience with search engines	yes,no	v_search_engine_experience
Graduation	Hochschulreife, Mittlere Reife, Hauptschulabschluss	v_graduation
Profession	<free text>	v_profession

Test protocol data were gathered by on-screen questionnaires before and after the experiment.

2.4 Experiment Control and Evaluation Database

Each experiment was fully supervised by a trained investigator. The investigator was with the test subjects throughout the experiment. The duration of one test including preparation, instruction and interview (without postprocessing) varies between 40 – 80 minutes. The postprocessing per experiment took about 160 minutes.
Except for the initial instruction and the postprocessing all required actions of an experiment are controlled by a Perl script which calls the different questionnaires, server startup and shutdown, saving of log data etc. automatically. To simplify field tests outside the campus all control scripts may be called from any workstation in our network or even remote over an secure shell Internet connection.
All data that are gathered during the process – manually or automatically – are inserted into the Evaluation Database via remote database calls or a database web interface.

A typical experiment within the formative evaluation required the following actions:

Preparation

- Investigator performs initial checks about server and DTAG services availability, connectivity of client (WLAN or UMTS), battery check, file space for logging.

Experiment

- Verbal instruction and handout of the task document to test subject
- Test subject fills out questionnaire A (only prior to task level 0)
- Investigator fills out test protocol questionnaire
- SW servers are started by remote secure shell and VNC; investigator checks for all SW modules working properly
- Investigator and test subject proceed to intended location
- Investigator checks for connectivity, starts SW client software and hands client and headset (if applicable) to test subject
- Test subject performs the tasks
- Investigator terminates SW client software
- Investigator and test subject return to office
- Investigator terminates SW server processes
- Log data are being saved
- Test subject fills out questionnaire B
- Test subject fills out questionnaire C (only after last test)
- Investigator handles IPR documents and payment to test subject (only after last test)

Postprocessing

- Investigator fills out test report sheet (comments, double check for important test conditions etc.)
- Client hardware is returned to lab and recharged
- Report sheet is verified to DB contents
- Recorded voice input is transcribed (see section 4)
- Interaction protocol is automatically extracted from log data (including voice input/output, presented media)
- Interaction protocol is manually segmented into interaction units and rated according a query success scheme (see section 4)
- Investigator issues money transfer as reimbursement for all performed tests (only after last test)

3 Test Subjects and Meta Data

A total of ten test subjects (3 male / 7 female) were carefully selected for the formative evaluation. The age ranges from 20 to 30. Professions and education levels are solely situated in the academic environment (9 test subjects were students and one test subject hold a university degree). All 10 test subjects declared to consult Internet services on a regular basis (8/10 daily); only one test subject reported to use dialogue systems once a week, the others only very rarely. All test subjects have had heard speech synthesis once in a while but do not use a device with speech output (such as a navigation system or SMS

reader) on a regular basis (values derived from questionnaire A, see appendix 8.7 for a complete listing of questionnaire results).
Aside from the person dependent test conditions as mentioned in section 2.3 the following meta data were recorded from each test subject:

Public data

Test subject code	[A-Z][A-Z][A-Z][A-Z]
Date of birth	YYYY-MM-DD
Gender	male, female
Mother tongue	<language code>
Mother tongue mother	<language code>
Mother tongue father	<language code>
Elementary School State	<state code>
Handedness	right, left, unknown
Project interest	yes, no
Comments	<free text>

Confidential data

Surname	<free text>
Name	<free text>
Street and number	<free text>
CIP code	<free text>
City	<free text>
Country	<country code>
Phone	<number>
Mobile phone	<number>
Fax	<number>
Email	<free text>

Care has been taken to separate public from confidential data. Only the database supervisor has access to the confidential part of the database.

4 Analysed Data

We distinguish three methodological different types of evaluation data: *technical, subjective* and *expert rated data*.

4.1 Technical Evaluation – Speech Recognition

As mentioned earlier all recorded voice input of the test subjects is being transcribed after the experiment by experienced phoneticians. The transcript is coded in a reduced subset of the SmartKom transliteration standard ([2]). Unnecessary tags were stripped from the transcript and different spellings of same words (e.g. 'Hauptbahnhof' vs. 'Haupt-Bahnhof'') were automatically equalized to avoid homophone errors in the results.
To handle the out-of-vocabulary (OOV) problem words not to be found in the recogniser's dictionary are replaced by the tag '<OOV>'; the latter was also done with detected out-of-vocabulary parts found in the output of the speech recogniser of SW. Thus, all out-of-vocabulary words in the transcript as well as in the ASR output can be matched on each other; special classes of OOV as being produced by the recogniser are not considered. Using a symbolic DP algorithm the resulting transcripts are automatically aligned to the corresponding output of the speech recognition engine of SW and the word correctness and word accuracy are calculated.[1]

4.2 Subjective Evaluation - Questionnaires

The subjective evaluation aims at a general judgement of the SmartWeb prototypes, judgement of certain aspects within the communication process (such as speech recognition, media presentation, ergonomics, speech output etc.) as well as to reveal positive or negative trends during the development of the prototypes.
For this purpose three different types of questionnaires were developed in close cooperation with SW partners[2]:

- **Questionnaire A** : general attitude and experience of test subjects towards dialogue systems, PDAs, search engines, speech synthesis.
- **Questionnaire B** : detailed rating of different aspects of a test session
- **Questionnaire C** : general attitude and experience of test subjects towards dialogue systems, PDAs, search engines; quality of speech synthesis in the SmartWeb system; general features of dialogue systems

Questionnaire A was answered by each test subject **prior to the first test session**. Aside from the speaker specific meta data as described in section 3 the test subjects were asked to rate their knowledge as well as their attitude towards automatic speech interfaces. Since five of the questions have been asked after the last performed test session, the found differences may indicate a change of attitude of the test subjects towards certain aspects of Human-Machine-Interfaces caused by the exposure to the SmartWeb prototypes (see Questionnaire C).
Technically the questionnaire was presented by a GUI that was automatically called prior to the first experiment; the entered data were inserted to the Evaluation DB by remote calls. See appendix 8.1 for a detailed description of the asked questions and appendix 8.7 for a complete listing of results.

Questionnaire B was answered by each test subject **right after each performed test session**. It comprises 29 ranking questions, two text fields to fill in positive or negative aspects of the system as well as a free text window to enter other comments. Each ranking question had an additional 'non-applicable' button and an optional comment field; ranking

[1] Correctness is defined as (N-D-S)/N while accuracy is defined as (N-D-S-I)/N where N = Number of words, S = number of substitutions, D = number of deletions and I = number of insertions.
[2] Mainly German DTAG Berlin and IMS Stuttgart

widths were either from 1 ... 6 (no neutral position) or 1...5 (neutral position at 3) between two extreme rankings, e.g. '*very bad 1 2 3 4 5 very good*'. Whenever applicable the extremities were arranged in such a way that the negative extreme is positioned to the left and the positive extreme was positioned to the right. The questionnaire form was generated automatically after the session was closed using the perl script *smwe.pl* published in [1] and the inserted values were directly transferred to the Evaluation Database.
See appendix 8.2 for a detailed listing of all 29 ranking questions.

Questionnaire C was answered by each test subject **right after the last test session**. It partly contains the same questions as Questionnaire A.
The aim of this questionnaire is to detect positive / negative trends in the overall judgement / attitude towards complex spoken language interfaces after being exposed to the SmartWeb system. Also, the overall quality of the speech synthesis of SmartWeb as well as general features of dialogue systems are evaluated in this questionnaire.

As with all subjective evaluations the problem of the reference arises: To which expectations do the test subjects compare the performance of the system. Partly this problem can be circumvented by the precise formulation of ratings; for instance:

Erroneous inputs could easily be corrected *I do not agree ... I agree*

Other, very general questions can only be analysed in relation to earlier / later test sessions, such as:

Using SmartWeb is ... *...boring ... ...fun*

Trends can be analysed across test sessions of a single test subject in time (questionnaire B) or across test sessions of all subjects regarding the prototype (questionnaire B) or with regard of the SmartWeb experience in general (questionnaire A vs. C).

4.3 Expert Evaluation of Query Success

Each response of the SmartWeb system depends on a variety of contributing modules interacting in a complex way. Some of these modules are themselves dependent of the state of the Internet itself. It is therefore – as with all systems with a higher complexity – not possible to assess the performance of each single module in isolation and sum up all performances to achieve an overall measurable performance of the system. This does of course not imply that single modules may not be assessed alone; for instance in the technical evaluation we apply this approach to the speech recognition engine, while in the subjective evaluation special assessments about the quality of the speech output are analysed. But for other, more interconnected modules such as the dialogue manager this approach is not feasible.

We therefore adopted the traditional *End-to-End evaluation* approach successfully applied earlier in the *Verbmobil* and *SmartKom* projects to come up with a measurable value representing the *query success* solely based on input and output of the system ('blackbox'). Since the input and output of the SW system is multimodal and furthermore different modalities may overlap in time, this query success cannot be extracted automatically from transcripts and log data:

In a first step a so called *transaction protocol* (TP) is automatically extracted from the available log data of a test session and time aligned with the transcript of the spoken input of the test subject.
The TP is then segmented into *interaction units* (IU) where each unit must contain at least either

- one user input and the corresponding system response,
- one user input without system response, or
- one system response without a prior user input

In a second pass each IU is labelled by an human expert with regard to *task type* (TT), *input modality* (IM) and *system answer* (SA).

Possible TT values are:

Code	Task Type
SYS	System greeting at start of new test session
POI	Point of interest: queries about touristic sites
OD	Open domain: queries outside of all SW domains
W	Weather: weather forecast in cities
C	Cinema: movie titles, actual programs, contents, ...
N	Navigation: requests about directions or route plans, maps
H	Hotel information
R	Restaurant information including price lists, reservations, bars, ...
F	Soccer: Games, players, world series, locations,...
E	Events: Cultural events such as concerts, theatre, ...
HC	Health care: directions to the next physician, hospital, pharmacy, ...
SPT	Public transportation: schedules, stations, bus lines, air traffic, ...
U	Unknown: functions that are not supported by SW, OffTalk, non–cooperative behaviour
+P	Picture(s) requested: including web cams, maps, etc.
+V	Video requested
+A	Audio requested
+T	Text(s) requested
+FQ	Follow–up question

For example:

Subject: "Show me a map of Munich with all the better restaurants on it."
TT: R+P
Subject: "Do you have any descriptions about these?"
TT: R+T+FQ

Possible *input modality* (IM) values are:

Code	Input Modality
S	Spoken input
T	Text input

Finally, the SA values describe the general success of the system response:

Code	System Answer
CO	Correct: SW answers the user query correctly
IC	Incorrect: SW presents an answer but the answer is not correct
PA	Partially correct: SW present several possible answers; the correct answer is among them but not at the first position
FA	Failed: SW ignores the input or recognizes the speech input but does not process it or presents the message *"Rückfrage"* and freezes
FU	Failed because of non-cooperative user behaviour, typos in text input, illogical queries
SN	SW processes the query but reports that no answer has been found

Note that the values of SA are *independent of the success of the speech recognition engine*, that is a perfectly recognized query may still lead to a SA value of 'FA' if the system does not process the query and does not issue a negative answer.

Example: Test subject types in: *"Picture of the Sears Tower"*
The system presents a picture of the Sears Tower.
Code: TT:OD+P; IM:T; SA:CO
Test subject says: *"Do you have a picture of the top as well?"*
The system presents a picture of a top-less model.
Code: TT:OD+P+FQ; IM:S; SA:IC

5 Summary of Results

5.1 Test Conditions

The table in appendix 8.5 gives an overview about the tested conditions across the 1+9 task levels of the formative evaluation. To summarize, the following table shows the total number of tests analysed per test condition:

Total	Environment	Connetivity	Headset	ASR activation
88	69 indoor	69 WLAN	85 MDA Pro	70 PTT
	19 outdoor	19 UMTS	3 none	18 open

5.2 Technical Evaluation

The speech recognition engine of SW was evaluated according to the method outlined in section 4.1.
The following table shows word correctness and accuracy in percent for all conducted experiments, across prototypes as well as the acoustical environment (indoor/outdoor) and ASR activation (push-to-talk/open microphone).[1]

Condition	all	0.5.1	0.6	0.8	indoor	outdoor	PTT	open
Correctness Accuracy	46.59 40.51	48.75 39.13	45.70 39.97	47.63 42.79	46.31 39.40	47.43 43.81	46.45 40.42	46.95 40.72
# of words	12199	1766	7563	2870	9129	3070	8663	3536

The measured *correctness* values do not show any significant improvement across prototypes nor differences between the other conditions.
The measured *accuracy* values show a significant improvement towards the prototype 0.8 which suggests that the speech engine **lowered the number of insertion errors**[2]. Also, it is remarkable that the accuracy is significantly **higher for outdoor environments** which is in contradiction to the general believe that speech recognition only works well in silent environments. The ASR activation seems to have no effect on the recognition results which also is in contradiction to the expectations that an open microphone will lead to worse speech input results.

The same values were determined for all tests conducted by a single test subject; **results across test subjects range from 39% to 51%** which confirms our assumption that no 'goat' was selected among the test subjects.

To see the influence of the task domain we calculated the same values for all test subjects across task types:

Task Level	Topics	Environment	Correctness / Accuracy	# of words
0	Training: usage of MDA Pro, listed queries	office	60.56 / 56.50	862
1	Watching soccer with friends, World Series 2006	office / cafe	42.10 / 31.89	3189
2	Planing a short visit to a German city	office / cafe	52.27 / 44.83	1209
3 + 5	Planing a visit to the cinema I + II	office / cafe	42.49 / 36.56	2768
4	Visit to Berlin	office	45.13 / 37.61	1582
6	List of pre-defined queries	office / street	47.38 / 45.16	1891
7	(Free selection of topic as in task level 1-4)	office / street	46.90 / 41.10	1467
8	Looking for a good physician	office / street	45.67 / 37.22	497

[1] 100% correctness means all spoken words have been recognized correctly but there might be additional inserted words that were not actually spoken; 100% accuracy means no additional inserted words. Consequently the correctness ranges from 0 to 100% while the accuracy might become negative (if many insertions take place).
[2] Correctness does not consider insertion errors while accuracy does.

Task Level	Topics	Environment	Correctness / Accuracy	# of words
9	Free queries, personal interests, typical search queries as given to Internet search engines	street	45.55 / 40.93	562

As it has to be expected the values for the training session 0 are significantly better than the remaining task levels, since test subjects used pre-fabricated queries in this task level. The remaining task levels are remarkable evenly distributed; tasks which the SW system was designed for (Soccer World Series 2006) did not score better than tasks with completely open topics (9). The best results in terms of correctness were achieved in the sight-seeing/navigation task (5) with 52,27%.

5.3 Subjective Evaluation – Questionnaires

5.3.1 Comments of Test Subjects Across Prototypes

In questionnaire B the test subjects as well as the investigator could record general remarks of the test subjects during or right after the test. These comments were classified in a scheme as given in the raw result reports in appendices 8.1-3.
Here we present only the most prominent findings (> 6%) across the three tested prototypes. The numbers represent the relative (and absolute in brackets) numbers of test subjects that uttered this kind of comment.
Unfortunately, no positive comments were reported that were counted in more than 6% of the tests; in fact only singular comments in each category were reported. Therefore we are not able to consider these reports as being statistically significant. See appendices 8.1-3 for a complete listing of all single positive comments (in German).

Negative Comments

Comment	0.5.1	0.6	0.8
"Found media are not presented"	39% (7)	8% (4)	*59% (–)*
"Processing is too slow"	28% (5)	10% (5)	*0% (0)*
"Results are not precise enough"	50% (9)	28% (14)	36% (8)
"Speech recognition insufficiant"	50% (9)	28% (14)	50% (11)
"Indication of processing state insufficiant"	17% (3)	28% (14)	36% (8)
"Insufficiant/non-functional ergonomics!"	17% (3)	20% (10)	41% (9)
"Usage of PTT / microphone indicator not intuitive"	17% (3)	26% (13)	45% (10)
"General Frustration caused by lack of success	11% (2)	28% (14)	32% (7)

Unfortunately there seems to be an increasing negative ranking of the SmartWeb systems with progressing prototype versions. It is inconclusive whether these trends are solely caused by a degrading performance of the SmartWeb prototypes or by effects introduced by the experimental setup such as an increasing frustration caused by persistent malfunctioning over an increasing number of tests.

5.3.2 Questionnaire B – Overall Judgment of SmartWeb

The following table summarizes the raw results of questionnaire B over all tests. A detailed description on the same data set together with the original question text in German is given in appendix 8.4.3, part *FRAGEBOGEN*.
The ranking scales of questionnaire B were mostly[1] arranged in a way that negative extremes are always to the left and positive to the right; therefore barycenters (indicated in underlined bold face) positioned in the right half of the table indicate positive ranking (11) and in the left half negative ranking (16).

	N.A.	1	2	3	4	5	6
Question: 1	0	1	6	13	21	**32**	15
Question: 2	0	0	0	4	8	**45**	**31**
Question: 3	14	5	19	13	**21**	16	0
Question: 4	0	3	7	18	**28**	**28**	4
Question: 5	1	7	19	10	**28**	15	8
Question: 6	2	2	20	16	**21**	20	7
Question: 7	15	18	**29**	10	7	9	0
Question: 8	0	40	**32**	14	2	0	0
Question: 9	0	26	**34**	14	11	3	0
Question: 10	6	0	1	7	16	**27**	**31**
Question: 11	0	**35**	**27**	7	5	11	3
Question: 12	1	2	13	8	15	**37**	12
Question: 13	5	9	**29**	**23**	21	1	-
Question: 14	5	3	14	**25**	**31**	10	-
Question: 15	6	**24**	**22**	**20**	16	0	-
Question: 16	8	5	11	**26**	**22**	16	-
Question: 17	5	10	**27**	**28**	17	1	-
Question: 18	1	**31**	**34**	10	3	9	0
Question: 19	0	**34**	**37**	7	4	6	0
Question: 20	2	22	**30**	12	15	5	2
Question: 21	0	9	14	5	11	12	**37**
Question: 22	0	4	5	8	10	**39**	**22**
Question: 23	1	**35**	**34**	13	2	3	0
Question: 24	0	28	**34**	25	1	0	-
Question: 25	1	**26**	**25**	18	9	9	0
Question: 26	0	0	0	4	9	24	**51**
Question: 27	0	6	5	13	**33**	**24**	7
Question: 28	0	21	**29**	20	11	6	1
Question: 29	0	9	6	**24**	**20**	15	14

It seems to be the case that the overall judgment of the Smartweb system was mostly average with a slight tendency to negative judgments.
Extreme outliers are:

- + Question 2 : "Readability of text output was good" (5 of 6)
- - Question 10 : "SmartWeb produced not enough output" (6 of 6)
- + Question 21 : "The dialogue flow was controlled by myself" (6 of 6)
- + Question 26 : "The combination of speech and text input makes sense" (6 of 6)

(variable parts of questions underlined)

5.3.3 Questionnaire B – Judgment Across Prototypes

To verify the influence of the SmartWeb prototype development on test subject judgements, we calculated the normalized barycenters for each question of questionnaire B across prototypes:

[1] Exceptions are questions 10 and 21 which were exempt from the barycenter counts.

<norm-barycenter> = sum [<question-rank> * <rank-count>] / <number-of-tests>

The following table shows the normalized barycenters for all questions except number 10 and 21 for the three tested SW prototypes. Questions with a positive trend are underlined. For your convenience we have included short English translations of the questions asked for all positive trends.

Prototype	**0.5.1**	**0.6**	**0.8**	
Question: 1	4.4	4.3	4.7	
Question: 2	5.3	5.2	5.0 *	
Question: 3	2.6	3.0	2.5	
Question: 4	3.7	3.9	4.2	"Structure of display is good"
Question: 5	3.5	3.5	3.5	
Question: 6	3.0	3.7	3.9	"SW always shows me when it's busy"
Question: 7	1.8	2.3	1.7 *	
Question: 8	1.9	1.6	1.9	
Question: 9	2.1	2.3	2.1	
Question: 11	1.5	2.4	2.6	"Everybody could use SW like I did"
Question: 12	4.3	4.2	4.2	
Question: 13	2.2	2.6	2.8	"Pleasant voice"
Question: 14	3.2	3.3	2.8 *	
Question: 15	2.2	2.2	2.1	
Question: 16	3.1	3.2	2.9 *	
Question: 17	2.2	2.6	2.5	
Question: 18	3.2	1.9	1.7 *	
Question: 19	1.9	1.9	2.1	
Question: 20	2.7	2.3	2.6	
Question: 22	3.8	4.8	4.8	"Always clear what to do next"
Question: 23	1.8	1.9	1.9	
Question: 24	2.1	1.9	2.1	
Question: 25	2.1	2.4	2.7	"Faster with speech input"
Question: 26	5.3	5.3	5.7	"Speech and text input makes sense"
Question: 27	3.8	3.9	4.3	"Feel comfortable with SW"
Question: 28	2.4	2.4	2.9	"Performance of SW is impressive"
Question: 29	3.9	3.7	4.0	

9 of 27 analysed questions show a significant improvement across prototypes. However, 5 of 27 shows the reverse effect (*), while the remaining 13 questions shows no conclusive trend. This slightly positive effect contradicts the findings in section 5.3.2. where negative comments increased with higher prototypes. One possible explanation is that test subjects got more adapted to the test situation and were more perceptive for errors in later experiments.

5.3.4 Questionnaire A vs. C – Trends before/after SmartWeb

General attitude/experience of the test subjects as being asked in the first section of questionnaire A[1] have already been discussed in section 3.
Appendices 8.7 and 8.8 show the raw results from questionnaires A and C respectively.

Before/After Effects
Five questions[2] were asked in questionnaire A and C to test for before / after effects caused by the exposure to the SmartWeb system. A comparison of the results reveal that only the answers to one question differ significantly:

[1] *Keys: experience_internet,experience_diasys, synthesis_**
[2] *Keys: payfor_diasys, opinion_service, opinion_human, opinion_personalassi, opinion_help*

"How much would you be willing to pay for a service that gives you unrestricted spoken access to the Web?"

Choices	Before SmartWeb	After SmartWeb
... nothing	3	5
... 10 Cents per minute	6	3
... 25 Cents per minute	1	2
... 50 Cents per minute	0	0
... 100 Cents per minute	0	0
... more than 200 Cents per minute	0	0

Speech Synthesis
Five questions[1] have been asked in questionnaire C regarding the quality / applicability / appropriateness of the speech output. The answers reflect a moderate acceptance of the speech output with exception of the question regarding the '*naturalness of the voice'* where the majority of the 10 test subjects opinioned for *'unnatural'.*

Special System Features
Six questions were asked in questionnaire C about certain general features of dialogue systems. Test subjects were instructed by the investigator not to give their opionion regarding the SmartWeb system but rather to dialogue systems in general. The motivation for these questions was to test wether the exposure to the SmartWeb system has heightened the awareness to certain ergonomic features.
In summary the test subjects only agreed to three important features:

- "It should always be possible to ask further questions about an already answered topic!" (9/10), e.g.:
  "*Tell me about the weather forecast in Berlin for today!" ... "What about tomorrow?"*
- "A dialogue system has always to notify the user that it is busy processing a query!" (10/10)
- "It must always be possible to correct/modify erroneous input or false speech recognition results from a previous query!" (10/10)

The majority of test subjects agreed that

- a dialogue system must acknowledge all queries right after their input
- a user should know beforehand which type of information can be retrieved using the dialogue system

Test subjects were undecided about whether the correct formulation of queries should be know beforehand to the user.

5.4 Expert Evaluation of Query Success

The log data and transcriptions of all 88 test sessions were processed and labelled according to the methodology given in section 4.3. In total 2102 interaction units were registered; from these 159 stemmed from the training session 0, which are considered separately in the following analysis.

[1] *Keys: uec_voice_fit_to_system, uec_voice_quality, uec_voice_pleasentness, uec_voice_naturalness, uec_voice_applicability*

General Query Success Rates

The following table shows the rating of the query success for all test sessions (codes explained in section 4.3):

Code	Task Levels 1–9		Task level 0	
	Count	%	Count	%
CO	123	6,33	29	18,23
IC	266	13,69	34	21,38
PA	21	1,08	6	3,77
SN	682	35,10	46	28,93
FA	338	17,40	20	12,58
FU	513	26,40	24	15,09
Total	1943	100,00	159	100,00

For an overall rating we could argue that SA types CO ('Answer correct'), PA ('Partially correct') and SN ('Processed query correctly but could not find information in the Web') should be considered as a successful task, while SA type IC ('Incorrect answer') and FA ('Failed to process query') should be considered as failures of the system. Finally, the SA type FU ('Failure because of uncooperative user behaviour') should not considered at all (for simplicity we present only the relative values in percent):

	Task Levels 1–9	Task level 0
Success (CO+PA+SN)	57,76%	60,00%
Failure (IC+FA)	42,24%	40,00%

The SmartWeb system processes every second input query correctly in the sense that it communicates a conclusive answer to the user.
Note that the number of cases where the system was unable to find an appropriate answer from the Web (SN) is with 35,1% very high compared to correct (6,33%) or partially correct (1,08%) answers.
The high percentage of user-induced failures 26,4% implies that there are still ergonomic problems with the SmartWeb concept, even after a longer exposure to the system.

Query Success vs. Input Modality

1627 IU (77,40%) have been labelled with Speech Input modality. To verify the impact of the speech recognition input we separated the query success rate as defined above according to the two different input modalities speech and text input:

Success (CO+PA+SN)	Task Levels 1–9	Task Level 0
Input Speech	53,28%	56,60%

Success (CO+PA+SN)	Task Levels 1–9	Task Level 0
Input Text	71,76%	72,41%

The text input modality yields a significantly higher query success rate for both, the training task level as well as the remaining task levels as it is to be expected, since the speech input engine often delivers wrong or incomplete input to the system.

Query Success vs. Task Type

Since the SmartWeb system handles different types of queries in a different manner, we analysed the query success rates as defined above for the different task types as defined in section 4.3. The following table shows the overall occurrence of a task type in the task levels 1-9 and the relative query success within all task types. Statistically valid values are set in bold face; the occurrence of the remaining task types was too low to yield reliable results.

Code	Task Type	Occurrence	Success (CO+PA+SN) Task Levels 1–9
POI	Point of interest: queries about touristic sites	**18,95%**	**64,21%**
OD	Open domain: queries outside of all SW domains	**22,31%**	**54,54%**
W	Weather: weather forecast in cities	3,36%	77,08%
C	Cinema: movie titles, actual programs, contents, ...	**23,7%**	**26,04%**
N	Navigation: requests about directions or route plans, maps	3,22%	78,26%
H	Hotel information	2,87%	65,85%
R	Restaurant information including price lists, reservations, bars, ...	**7,41%**	**42,45%**
F	Soccer: Games, players, world series, locations,...	**9,65%**	**67,39%**
E	Events: Cultural events such as concerts, theatre, ...	2,17%	25,80%
HC	Health care: directions to the next physician, hospital, pharmacy, ...	**5,59%**	**52,50%**
SPT	Public transportation: schedules, stations, bus lines, air traffic, ...	1,82%	57,69%

Code	Task Type	Occurrence	Success (CO+PA+SN) Task Levels 1–9
U[1]	Unknown: functions that are not supported by SW, Off–Talk, non–cooperative behaviour	–	–

The query success appears to be quite uniform across the SmartWeb task types. Please note that the main contribution to the query success as shown in the table is due to the rating SN ('Processed query correctly but could not find information in the Web'). Correct (CO) or partially correct answers (PA) were very sparse in all task types.

[1] Task type 'Unknown' was in the majority classified with SA = FU, that is a 'user induced failure' and can therefore not be considered for the query success as defined above.

6 Conclusion

The technical evaluation of the speech input recognition with respect to word correctness yields an average value of 46,6% without any significant improvement across prototypes nor differences between the other test conditions.
The measured *accuracy* is about 40,5% and shows a significant improvement towards the prototype 0.8 which suggests that the speech engine was redesigned to lower the number of insertion errors.
Word correctness is remarkable evenly distributed across different task domains; tasks which the SW system was designed for (e.g. Soccer World Series 2006) did not score better than tasks with completely open topics. The best results in terms of word correctness were achieved in the sight-seeing/navigation task with 52,27%.

Considering the fact that we conducted a field trial under realistic conditions (outdoors, heavy noise, real network) and the fact that the system is able to process incomplete or malformed input queries this is not a bad result.

The analysis of the test subjects' comments shows an increasing negative ranking of the SmartWeb systems with progressing prototype versions. It is inconclusive whether these trends are solely caused by a degrading performance of the SmartWeb prototypes or by effects introduced by the experimental setup such as an increasing frustration caused by persistent malfunctioning. These results contradict the results obtained from the test subjects questionnaires (see below).

The analysis of the test subjects' questionnaire shows that the overall judgment of the Smartweb system was mostly average with a slight tendency to negative judgments. 9 of 27 analysed questions show a significant improvement across prototypes. However, 5 of 27 shows the reverse effect, while the remaining 13 questions show no conclusive trend.

The expert rating of 2103 interaction units with regard to the query success shows that the SmartWeb system processes 57% of well-formed input queries correctly in the sense that it communicates a conclusive answer to the user. Note that the number of cases where the system was nevertheless unable to find an appropriate answer from the Web is with 35,1% very high compared to correct (6,33%) or partially correct (1,08%) answers. Query success rates are significantly higher for text input modality than speech input modality which has to be expected. However, 77% of input queries were in fact performed in speech which implies that test subjects nevertheless preferred spoken input over the more reliable but tedious text input modality.

7 References

[1] Chin, J.P., Diehl, V.A., Norman, K.L. (1988) Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. ACM CHI'88 Proceedings, pp. 213-218.
[2] D. Oppermann, S. Burger, S. Rabold, N. Beringer: Transliteration spontansprachlicher Daten – Transliterationslexikon SmartKom. SmartKom TechDok 2 – 2001.

8 Appendix

8.1 Questionnaire A

The following questions have been asked:

Key: uea_experience_internet
Question=**"Ich rufe Informationen aus dem Internet ab und zwar..."**
Pulldown: **mehrmals am Tag**
etwa einmal pro Tag
ein paar Mal pro Woche
höchstens ein paar Mal im Monat
praktisch nie

Key: uea_experience_diasys
Question=**"Wie häufig verwenden Sie automatische Dialogsysteme? Z.B. telefonische Banksysteme, Autosteuerung per Sprache, autom. Auskunft"**
Pulldown: **Ich entwickele selber Dialogsysteme**
Ich verwende sie jeden Tag
Ich verwende sie jede Woche
Ich verwende sie nicht mehr als einmal im Monat
Ich verwende nur selten ein Dialogsystem
Ich habe noch nie ein Dialogsystem benutzt

Key: uea_pre_opinion_diasys
Question=**"Angenommen es gibt ein perfekt funktionierendes Auskunftsystem, dann spreche ich..."**
Pulldown: **1 ... trotzdem lieber mit einem Menschen**
2
3
4
5 ... lieber mit der Maschine

Key: uea_pre_payfor_diasys
Question=**"Für einen Service, der mir mit alltäglicher Sprache unbegrenzten Zugang zu Web-Inhalten bietet, wäre ich ..."**
Pulldown: **nicht bereit, etwas zu zahlen**
bereit, 10 Cent pro Minute zu zahlen
bereit, 25 Cent pro Minute zu zahlen
bereit, 50 Cent pro Minute zu zahlen
bereit, 1 Euro pro Minute zu zahlen
mehr als 2 Euro pro Minute zu zahlen

Key: uea_pre_opinion_service
Question=**"Ein intelligentes Dialogsystem kann mir niemals den gleichen Service bieten wie ein Mensch."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch

Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uea_pre_opinion_human
Question=**"Ich lege neben der reinen Information auch großen Wert auf die menschliche Seite der Kommunkation."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uea_pre_opinion_personalassi
Question=**"Ich hätte gerne einen persönlichen Assistenten (Mensch oder Maschine), mit dem ich zu jeder Zeit auf natürliche Weise auf Inhalte des Webs zugreifen kann."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uea_synthesis_simple
Question=**"Wie oft haben Sie schon eine einfache automatische Telefonansage gehört?"**
Pulldown: **sehr oft**
oft
einige Male
selten
nie

Key: uea_synthesis_sms
Question=**"Wie oft haben Sie schon einen SMS-Vorlese-Service gehört?"**
Pulldown: **sehr oft**
oft
einige Male
selten
nie

Key: uea_synthesis_nav_simple
Question=**"Wie oft haben Sie schon die Stimme eines Nav.systems ohne Ansage von Straßen- und Ortsnamen gehört?"**
Pulldown: **sehr oft**
oft
einige Male
selten
nie

Key: uea_synthesis_nav_complex

Question=**"Wie oft haben Sie schon die Stimme eines Nav.systems mit Ansage von Strassen- und Ortsnamen gehört?"**
Pulldown: **sehr oft**
oft
einige Male
selten
nie

Key: uea_pre_opinion_help
Question=**"Ein Auskunftsystem muss in der Lage sein, dem Benutzer zu helfen, wie er am schnellsten an die gesuchte Information kommen kann."**
Pulldown: **Finde ich extrem wichtig**
Ziemlich wichtig
Tendenziell wichtig
Tendenziell unwichtig
Eher unwichtig
Halte ich für völlig irrelevant

8.2 Questionnaire B

A screen shot of the Web form is shown in fig. A1.
The following ranking questions have been asked (ranking width in brackets):

1 Die Bedienung von SmartWeb ist ...
schwierig ... leicht (6)
2 Die Lesbarkeit der Schrift war...
sehr schlecht ... sehr gut (6)
3 Die Hervorhebungen erleichterten die Bedienung.
stimme ich nicht zu ... stimme ich zu (6)
4 Die Anordnung der Informationen auf dem Display finde ich...
verwirrend ... übersichtlich (6)
5 Die Ausdrucksweise von SmartWeb war...
inkonsistent ... konsistent (6)
6 SmartWeb zeigt mir, wenn es beschäftigt ist.
nie ... immer (6)
7 Die FehlermeldAuswertung_FBb_vp5813.outungen von SmartWeb sind...
wenig hilfreich ... sehr hilfreich (6)
8 Die Geschwindigkeit von SmartWeb fand ich...
zu langsam ... zu schnell (5)
9 Das SmartWeb funktionierte...
unzuverlässig ... reibungslos (6)
10 SmartWeb lieferte...
zu viele Informationen ... zu wenig Informationen (6)
11 So wie SmartWeb in der heutigen Sitzung funktioniert hat, könnte auch jeder andere damit umgehen.
stimme ich nicht zu ... stimme ich zu (6)
12 Fehleingaben konnte ich leicht korrigieren.
stimme ich nicht zu ... stimme ich zu (6)
13 Wie angenehm fanden Sie die Stimme?
sehr unangenehm ... sehr angenehm (5)

14 Welche Anstrengung war nötig, um die Äußerungen zu verstehen?
selbst größte Anstrengung reicht nicht zum Verstehen ... es war keine Anstrengung zum Verstehen erforderlich (5)
15 Wie würden Sie die Natürlichkeit der Stimme einschätzen?
sehr unnatürlich ... sehr natürlich (5)
16 Fanden Sie heute bestimmte Wörter schwer zu verstehen?
ständig ... nie (5)
17 Wie würden Sie insgesamt die Sprachqualität der gehörten Äußerungen beurteilen?
sehr schlecht ... sehr gut (5)
18 Die Aktivierung der Spracheingabe fand ich ...
einfach ... umständlich (6)
19 SmartWeb hat mich schnell zur gewünschten Information geführt.
trifft nicht zu trifft zu (6)
20 SmartWeb hat mir die richtigen Informationen geliefert.
trifft nicht zu trifft zu (6)
21 Der Gesprächsverlauf wurde eher von...
SmartWeb bestimmt ... mir selber bestimmt (6)
22 Ich wusste immer, wie ich die nächste Eingabe (per Sprache, Tastatur, Stift) machen konnte.
trifft nicht zu ... trifft zu (6)
23 Ich musste meine Eingaben wiederholen.
praktisch jedes mal ... nie (6)
24 Die Pausen zwischen Eingabe und Antwort erschienen mir ...
sehr lang ... sehr kurz (5)
25 Mit Hilfe der Spracheingabe komme ich schneller ans Ziel
trifft nicht zu ... trifft zu (6)
26 Die Kombination von Stift- und Spracheingabe finde ich sinnvoll
trifft nicht zu ... trifft zu (6)
27 Im Gespräch mit SmartWeb fühlte ich mich ...
unwohl ... wohl (6)
28 Ich bin von der Leistung von SmartWeb ...
enttäuscht ... beeindruckt (6)
29 Der Umgang mit SmartWeb hat ...
mich gelangweilt ... mir Spaß gemacht (6)

SmartWeb Evaluation Questionnaire

Mit diesem Fragebogen beurteilen Sie die soeben erfolgte Benutzung des SmartWeb Systems. Das Ausfüllen dauert etwa 15 Minuten.

- Bitte beantworten Sie unbedingt alle Fragen.
- Bei Fragen, die Ihrer Meinung nach nicht auf das soeben durchgeführte Experiment passen, wählen Sie bitte den Knopf: **NA**
- Sie können bei Bedarf einzelne Fragen durch Klicken auf das Icon kommentieren
- Erst wenn Sie alle Fragen beantwortet und etwaige Kommentare abgegeben haben, klicken Sie zum Abschluss auf: **Daten sichern**

Sprecherkürzel: **Nachname:** **Session-ID:**

Insgesamte Beurteilung								**NA**
1. Die Bedienung von SmartWeb ist ...	schwierig	○	○	○	○	○	○ leicht	○
Der Bildschirm von SmartWeb								**NA**
2. Die Lesbarkeit der Schrift war...	sehr schlecht	○	○	○	○	○	○ sehr gut	○
3. Die Hervorhebungen erleichterten die Bedienung.	stimme ich nicht zu	○	○	○	○	○	○ stimme ich zu	○
4. Die Anordnung der Informationen auf dem Display finde ich...	verwirrend	○	○	○	○	○	○ übersichtlich	○
Ausgaben von SmartWeb								**NA**
5. Die Ausdrucksweise von SmartWeb war...	inkonsistent	○	○	○	○	○	○ konsistent	○
6. SmartWeb zeigt mir, wenn es beschäftigt ist.	nie	○	○	○	○	○	○ immer	○
7. Die Fehlermeldungen von SmartWeb sind...	wenig hilfreich	○	○	○	○	○	○ sehr hilfreich	○
Fähigkeiten von SmartWeb								**NA**
8. Die Geschwindigkeit von SmartWeb fand ich...	zu langsam	○	○	○	○	○	○ zu schnell	○
9. Das SmartWeb funktionierte...	unzuverlässig	○	○	○	○	○	○ reibungslos	○
10. SmartWeb lieferte...	zu viele Informationen	○	○	○	○	○	○ zu wenig Informationen	○
11. So wie SmartWeb in der heutigen Sitzung funktioniert hat, könnte auch jeder andere damit umgehen.	stimme ich nicht zu	○	○	○	○	○	○ stimme ich zu	○
12. Fehleingaben konnte ich leicht korrigieren.	stimme ich nicht zu	○	○	○	○	○	○ stimme ich zu	○
Die folgenden Fragen beziehen sich auf die synthetische Stimme in der *heutigen Sitzung*, nicht auf frühere Sitzungen								**NA**
13. Wie angenehm fanden Sie die Stimme?	sehr unangenehm	○	○	○	○	○	○ sehr angenehm	○
14. Welche Anstrengung war nötig, um die Äußerungen zu verstehen?	selbst größte Anstrengung reicht nicht zum Verstehen	○	○	○	○	○	○ es war keine Anstrengung zum Verstehen erforderlich	○
15. Wie würden Sie die Natürlichkeit der Stimme einschätzen?	sehr unnatürlich	○	○	○	○	○	○ sehr natürlich	○

Fig. A1 : Screen shot of questionnaire B

8.3 Questionnaire C

The following questions have been asked:

Key: uec_post_payfor_diasys
Question: **"Für einen Service, der mir jederzeit auf natürliche Weise unbegrenzten Zugang zu Web-Inhalten bietet, wäre ich ..."**
Pulldown: **nicht bereit, etwas zu zahlen**
bereit, 10 Cent pro Minute zu zahlen
bereit, 25 Cent pro Minute zu zahlen
bereit, 50 Cent pro Minute zu zahlen
bereit, 1 Euro pro Minute zu zahlen
mehr als 2 Euro pro Minute zu zahlen

Key: uec_post_opinion_service
Question: **"Ein intelligentes Dialogsystem kann mir niemals den gleichen Service bieten wie ein Mensch."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig

Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uec_post_opinion_human
Question: **"Ich lege neben der reinen Information auch großen Wert auf die menschliche Seite der Kommunkation."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uec_post_opinion_personalassi
Question: **"Ich hätte gerne einen persönlichen Assistenten, mit dem ich zu jeder Zeit auf natürliche Weise auf Inhalte des Webs zugreifen kann."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uec_post_opinion_help
Question: **"Ein Auskunftsystem muss in der Lage sein, dem Benutzer zu helfen, wie er am schnellsten an die gesuchte Information kommen kann."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uec_voice_fit_to_system
Question: **"Wie gut passt die Stimme zum getesteten System?"**
Pulldown: **Sehr gut**
Gut
Ordentlich
Schlecht
Sehr schlecht

Key: uec_voice_quality
Question: **"Wie würden Sie nun, nachdem Sie mehrere Sitzungen mit SmartWeb absolviert haben, insgesamt die Sprachqualität der Systemäußerungen beurteilen?"**
Pulldown: **Sehr gut**
Gut
Ordentlich
Schlecht
Sehr schlecht

Key: uec_voice_pleasentness
Question: **"Wie angenehm fanden Sie die Stimme insgesamt, in allen Sitzungen?"**
Pulldown: **Sehr angenehm**
Angenehm
Neutral
Unangenehm
Sehr unangenehm

Key: uec_voice_naturalness
Question: **"Wie würden Sie die Natürlichkeit der Stimme insgesamt, in allen Sitzungen, einschätzen?"**
Pulldown: **Sehr natürlich**
Natürlich
Neutral
Unnatürlich
Sehr unnatürlich

Key: uec_voice_applicability
Question: **"Finden Sie, dass die Stimme im SmartWeb-System eingesetzt werden kann?"**
Pulldown: **Ja**
Nein

Anleitung für Versuchsleiter:
Ab hier beziehen sich die Fragen bzw. Aussagen ganz allgemein auf Intelligente Dialogsysteme und nicht notwendigerweise auf SmartWeb D.h. es soll nicht geprüft werden, ob SmartWeb diese Egenschaft hat, sondern, ob so eine Eigenschaft in einem System wie SmartWeb wünschenswert wäre oder nicht.

Key: uec_additional_questions
Question: **"Es sollte möglich sein, zu einer erhaltenen Information Zusatzfragen stellen zu können."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uec_defined_formulation
Question: **"Man sollte von Anfang an wissen, wie man seine Fragen formulieren soll."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uec_know_infotype

Question: **"Man sollte von Anfang an wissen, welche Art von Informationen ein System liefern kann."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uec_acknowledge
Question: **"Ein gut funktionierendes System sollte jede meiner Eingaben bestätigen."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uec_processing
Question: **"Ein gut funktionierendes System zeigt mir an, wenn es beschäftigt ist."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

Key: uec_modify_input
Question: **"Falsche Eingabe oder falsch erkannte Eingaben sollten leicht korrigiert werden können."**
Pulldown: **Stimme ich voll zu**
Weitgehend richtig
Tendenziell richtig
Tendenziell falsch
Weitgehend falsch
Stimme ich überhaupt nicht zu

8.4 Raw Results of Formative Evaluation

The following 3 report listings contains the overall results found in the formative evaluations as being published during the project phase Nov 2006 – June 2007.
The reports cover roughly the results obtained testing the prototypes 0.5.1, 0.6 and 0.8.[1]

[1] The data reported here contain in fact the results of more than the official 10 test subjects, since some test subjects were not able to finish the complete evaluation test series and were therefore later excluded from the evaluation. This refers to the speakers *AAJJ*

Report	*Duration*	*Prototype*	*# test subjects*	*# sessions*
1	08.11. - 31.01.2007	0.5.1	8	18
2	05.02. – 18.04.2007	0.6	10	50
3	19.04 – 18.07.2007	0.8	8	22

Please note that here all reported comments and questionnaire answers are given without any evaluations regarding their consequences as well as their validity. Please refer to chapter 5 for a discussion and analysis based on these data.

8.4.1 Formative Evaluation - Report 1

```
SmartWeb Evaluation

06.02.2007     Florian Schiel, Hannes Mögele

Version Prototyp:  0.5.1  (Subversion 6520)
Endgerät:          MDA Pro
Server:            2 x ( 2,8GHz, 3GB RAM, SuSE 9.2 )

Einstellung am SW-Client:
- Mikrophon muss für jede Anfrage freigeschaltet werden
- MDA Pro Headset (kein Freisprechen)
- GPS-Lokalisation: Berlin


Begleitende Evaluation - Report 1 (08.11.2006 - 30.01.2007)

Dieser Report berichtet über Mängel beim Betrieb des SmartWeb-Prototypen, die
während der begleitenden Evaluation in den bis zum Berichtszeitpunkt
durchgeführten Evaluationsexperimenten aufgetreten sind. Mängel-Reports der
begleitenden Evaluation sind mit 'Bii.x' (ii = Nummer der Kategorie,
x = fortlaufende Nummer der Subkategorie) bezeichnet.
Bitte beziehen Sie sich in der Korrespondenz auf diesen Bezeichner.

Als Quellen für diesen Report wurden
(1) die obligatorischen Bewertungen der Sessions durch die jeweiligen
    Versuchpersonen (vgl. hierzu den Abschnitt 'Fragebogen'),
(2) die freien Kommentare der Versuchspersonen (fakultativ) und
(3) die Beobachtungen der Versuchsleitung
herangezogen.

(2) beinhaltet Anmerkungen der Versuchspersonen, die diese zusätzlich zum
Fragebogen für besonders erwähnenswert hielten; deshalb werden auch Anmerkungen
berichtet, die nur in einer Session und von einer Versuchsperson geäußert wurden.

Aus (2) und (3) wurden bisher folgende Klassen gebildet:

B01 Medientypen
B02 Performanz
B03 Spracherkennung
B04 Sprachsynthese
B05 GUI
B06 Motivation / Emotion
B07 Semantische Analyse


Für diesen Report wurden 18 Evaluationsexperimente mit 8 Versuchspersonen
brücksichtigt. Davon waren 8 erste, 7 zweite und 3 dritte Sitzungen.

==================================
 Sprecher | Session | Aufnahmedatum
==================================
```

```
 AAAJ     | e005    | 2006-11-16
 AAAJ     | e023    | 2007-01-29
-----------------------------------
 AAAM     | e006    | 2006-11-23
 AAAM     | e014    | 2007-01-10
 AAAM     | e025    | 2007-01-30
-----------------------------------
 AAAT     | e022    | 2007-01-25
 AAAT     | e009    | 2006-12-08
-----------------------------------
 AAAW     | e021    | 2007-01-25
 AAAW     | e018    | 2007-01-23
-----------------------------------
 AADA     | e020    | 2007-01-24
 AADA     | e015    | 2007-01-11
 AADA     | e017    | 2007-01-17
-----------------------------------
 AADD     | e019    | 2007-01-24
 AADD     | e016    | 2007-01-17
 AADD     | e024    | 2007-01-30
-----------------------------------
 AAGW     | e002    | 2007-01-24
 AAGW     | e004    | 2006-11-14
-----------------------------------
 AAJJ     | e001    | 2006-11-08
===================================
```

Die Versuchspersonen merkten folgende Punkte als besonders positiv an:

* die Idee von Smartweb
* die multimodalen Eingabemöglichkeiten (vgl. Frage 26)
* die Übersichtlichkeit des GUI und das intuitive Design
* die Möglichkeit sich unterschiedliche Medientypen anzeigen zu lassen
* Freude im Umgang mit dem System (vgl. Frage 29)
* die Funktionalität des Wetterberichts
* vgl. auch Frage 1, 2 und 21 des Fragebogens

MÄNGEL

```
================================================================================
B01.1 [Medientypen] fehlende Anzeige der gefundenen Medientypen
--------------------------------------------------------------------------------
in 7 Sessions von 5 Versuchspersonen
--------------------------------------------------------------------------------

================================================================================
B02.1 [Performanz] zu langsame Verarbeitungsgeschwindigkeit (vgl. Frage 8)
--------------------------------------------------------------------------------
in 5 Sessions von 4 Versuchspersonen
--------------------------------------------------------------------------------

================================================================================
B02.2 [Performanz] ungenügende Bearbeitungsgenauigkeit
--------------------------------------------------------------------------------
* Fragen zum Thema Fußball wurden zwar erkannt aber nie richtig beantwortet
* auf die Frage nach dem WM-Sieger zwischen 1954 und 1970 wird ein Video eines
  Fußballspiels geliefert - alle Antworten sind aber falsch
* trotz korrekter Spracherkennung und semantische Analyse wird keine Antwort
  generiert
* Zahlenangaben sind sehr ungenau / unzuverlässig
* es wird nicht zwischen Bundeskanzler und -präsident unterschieden
--------------------------------------------------------------------------------
in 9 Sessions von 7 Versuchspersonen
--------------------------------------------------------------------------------

================================================================================
B03.1 [Spracherkennung] ungenügende Spracherkennerleistung
--------------------------------------------------------------------------------
```

```
* (fremdsprachige) Eigennamen werden selten erkannt
* die Spracherkennung erkennt das Wort "Papst" nicht
* bei erhöhtem Geräuschpegel keine Spracherkennung
* statt "Berlin" wird sehr oft "Wendelin" oder "Belgien" erkannt
--------------------------------------------------------------------------------
in 9 Sessions von 6 Versuchspersonen
--------------------------------------------------------------------------------

================================================================================
B04.1 [Sprachsynthese] unflüssige Sprachsynthese
--------------------------------------------------------------------------------
* "stottert"
* "gebrochen"
* "so ungefähr, als ob er Parkinson hï¿½tte"
--------------------------------------------------------------------------------
in 4 Sessions von 3 Versuchspersonen
--------------------------------------------------------------------------------

================================================================================
B04.2 [Sprachsynthese] schlecht verständliche Sprachsynthese
--------------------------------------------------------------------------------
in 1 Session von 1 Versuchsperson
--------------------------------------------------------------------------------

================================================================================
B04.3 [Sprachsynthese] fehlende Sprachsynthese
--------------------------------------------------------------------------------
in 1 Session von 1 Versuchsperson
--------------------------------------------------------------------------------

================================================================================
B04.4 [Sprachsynthese] Persona
--------------------------------------------------------------------------------
* "Der kommt mir so dumm vor."
* "Da wird man ja richtig aggressiv"
--------------------------------------------------------------------------------
in 1 Session von 1 Versuchsperson
--------------------------------------------------------------------------------

================================================================================
B05.1 [GUI] nicht intuitiv
--------------------------------------------------------------------------------
in 1 Session von 1 Versuchsperson
--------------------------------------------------------------------------------

================================================================================
B05.2 [GUI] Anzeige des Bearbeitungsstatus'
--------------------------------------------------------------------------------
* unzureichende Informationen über den Fortschritt oder Abbruch eines Vorgangs
--------------------------------------------------------------------------------
in 3 Sessions von 3 Versuchspersonen
--------------------------------------------------------------------------------

================================================================================
B05.3 [GUI] Bedienung
--------------------------------------------------------------------------------
* Suchergebnisse nicht klickbar
* fehlende Abbruchtaste
--------------------------------------------------------------------------------
in 3 Sessions von 3 Versuchspersonen
--------------------------------------------------------------------------------

================================================================================
B05.4 [GUI] Mikrofonsymbol
--------------------------------------------------------------------------------
* trotz Instruktion sind die VPs immer wieder irritiert, wann die
  Spracherekennung aktiviert ist und wann nicht
* Symbolik (durchgestrichen/grün/rot) des Spracherkennungsstatus ist verwirrend
--------------------------------------------------------------------------------
in 3 Sessions von 3 Versuchspersonen
--------------------------------------------------------------------------------
```

```
================================================================================
B06.1 [Motivation / Emotion] Frustration aufgrund von B02.2, B03.1
--------------------------------------------------------------------------------
in 2 Sessions von 2 Versuchspersonen
--------------------------------------------------------------------------------

================================================================================
B07.1 [Semantische Analyse] verwirrende Ausgabe der semantischen Analyse
--------------------------------------------------------------------------------
in 1 Session von 1 Versuchsperson
--------------------------------------------------------------------------------



### FRAGEBOGEN ###############################################################

Es wird jeweils die laufende Nummer der Fragen (Frage 1), das Item
('Die Bedienung von SmartWeb ist ...') mit Attributen ('schwierig' - 'leicht')
gegeben. Die Ziffer zwischen den Attributen gibt die Skalierung ('[5]' oder '[6]')
wieder.


================================================================================
Frage 1
--------------------------------------------------------------------------------
Die Bedienung von SmartWeb ist ...
schwierig [6] leicht
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 6 | 10| 0 |
-------------------------
================================================================================

================================================================================
Frage 2
--------------------------------------------------------------------------------
Die Lesbarkeit der Schrift war...
sehr schlecht [6] sehr gut
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 2 | 8 | 8 |
-------------------------
================================================================================

================================================================================
Frage 3
--------------------------------------------------------------------------------
Die Hervorhebungen erleichterten die Bedienung.
stimme ich nicht zu [6] stimme ich zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 1 | 6 | 1 | 2 | 5 | 0 |
-------------------------
keine Bewertung: 3
================================================================================

================================================================================
Frage 4
--------------------------------------------------------------------------------
Die Anordnung der Informationen auf dem Display finde ich...
verwirrend [6] übersichtlich
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 0 | 4 | 4 | 4 | 4 | 2 |
-------------------------
================================================================================
```

```
================================================================================
Frage 5
--------------------------------------------------------------------------------
Die Ausdrucksweise von SmartWeb war...
inkonsistent [6] konsistent
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 1 | 3 | 2 | 5 | 6 | 0 |
-------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 6
--------------------------------------------------------------------------------
SmartWeb zeigt mir, wenn es beschäftigt ist.
nie [6] immer
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 2 | 3 | 5 | 4 | 2 | 1 |
-------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 7
--------------------------------------------------------------------------------
Die Fehlermeldungen von SmartWeb sind...
wenig hilfreich [6] sehr hilfreich
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 5 | 11| 0 | 0 | 1 | 0 |
-------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 8
--------------------------------------------------------------------------------
Die Geschwindigkeit von SmartWeb fand ich...
zu langsam [5] genau richtig
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 6 | 7 | 4 | 1 | 0 |
--------------------
================================================================================

================================================================================
Frage 9
--------------------------------------------------------------------------------
Das SmartWeb funktionierte...
unzuverlässig [6] reibungslos
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 6 | 7 | 3 | 2 | 0 | 0 |
-------------------------
================================================================================

================================================================================
Frage 10
--------------------------------------------------------------------------------
SmartWeb lieferte...
zu viele Informationen [6] zu wenig Informationen
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
```

```
|---|---|---|---|---|---|
| 0 | 0 | 2 | 2 | 8 | 4 |
-------------------------
keine Bewertung: 2
================================================================================

================================================================================
Frage 11
--------------------------------------------------------------------------------
So wie SmartWeb in der heutigen Sitzung funktioniert hat, könnte auch jeder
andere damit umgehen.
stimme ich nicht zu [6] stimme ich zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 13| 4 | 0 | 0 | 0 | 1 |
-------------------------
================================================================================

================================================================================
Frage 12
--------------------------------------------------------------------------------
Fehleingaben konnte ich leicht korrigieren.
stimme ich nicht zu [6] stimme ich zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 0 | 2 | 3 | 3 | 7 | 3 |
-------------------------
================================================================================

================================================================================
Frage 13
--------------------------------------------------------------------------------
Wie angenehm fanden Sie die Stimme?
sehr unangenehm [5] sehr angenehm
---------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1 | 6 | 8 | 1 | 0 |
---------------------
keine Bewertung: 2
================================================================================

================================================================================
Frage 14
--------------------------------------------------------------------------------
Welche Anstrengung war nötig, um die Äußerungen zu verstehen?
selbst größte Anstrengung reicht nicht zum Verstehen [5] es war keine
Anstrengung zum Verstehen erforderlich
---------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 0 | 3 | 5 | 5 | 3 |
---------------------
keine Bewertung: 2
================================================================================

================================================================================
Frage 15
--------------------------------------------------------------------------------
Wie würden Sie die Natürlichkeit der Stimme einschätzen?
sehr unnatürlich [5] sehr natürlich
---------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 4 | 4 | 4 | 4 | 0 |
---------------------
keine Bewertung: 2
================================================================================
```

```
================================================================================
Frage 16
--------------------------------------------------------------------------------
Fanden Sie heute bestimmte Wörter schwer zu verstehen?
ständig [5] nie
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1 | 4 | 3 | 4 | 4 |
--------------------
keine Bewertung: 2
================================================================================

================================================================================
Frage 17
--------------------------------------------------------------------------------
Wie würden Sie insgesamt die Sprachqualität der gehörten Äußerungen beurteilen?
sehr schlecht [5] sehr gut
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 3 | 5 | 6 | 2 | 0 |
--------------------
keine Bewertung: 2
================================================================================

================================================================================
Frage 18
--------------------------------------------------------------------------------
SmartWeb konnte meine Fragen beantworten.
keine einzige [6] praktisch alle
------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 3 | 3 | 5 | 2 | 5 | 0 |
------------------------
================================================================================

================================================================================
Frage 19
--------------------------------------------------------------------------------
SmartWeb hat mich schnell zur gewünschten Information geführt.
trifft nicht zu [6] trifft zu
------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 7 | 8 | 1 | 1 | 1 | 0 |
------------------------
================================================================================

================================================================================
Frage 20
--------------------------------------------------------------------------------
SmartWeb hat mir die richtigen Informationen geliefert.
trifft nicht zu [6] trifft zu
------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 4 | 6 | 3 | 3 | 1 | 1 |
------------------------
================================================================================

================================================================================
Frage 21
--------------------------------------------------------------------------------
Der Gesprächsverlauf wurde eher von...
SmartWeb bestimmt [6] mir selber bestimmt
------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 2 | 1 | 1 | 1 | 5 | 8 |
```

```
------------------------
================================================================================

================================================================================
Frage 22
--------------------------------------------------------------------------------
Ich wusste immer, wie ich die nächste Eingabe (per Sprache, Tastatur, Stift)
machen konnte.
trifft nicht zu [6] trifft zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 2 | 3 | 1 | 4 | 6 | 2 |
------------------------
================================================================================

================================================================================
Frage 23
--------------------------------------------------------------------------------
Ich musste meine Eingaben wiederholen.
praktisch jedes mal [6] nie
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 7 | 6 | 2 | 1 | 1 | 0 |
------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 24
--------------------------------------------------------------------------------
Die Pausen zwischen Eingabe und Antwort erschienen mir ...
... sehr lang [5] ... sehr kurz
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 5 | 6 | 6 | 1 | 0 |
--------------------
================================================================================

================================================================================
Frage 25
--------------------------------------------------------------------------------
Mit Hilfe der Spracheingabe komme ich schneller ans Ziel
trifft nicht zu [6] trifft zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 5 | 8 | 3 | 0 | 1 | 1 |
------------------------
================================================================================

================================================================================
Frage 26
--------------------------------------------------------------------------------
Die Kombination von Stift- und Spracheingabe finde ich sinnvoll.
trifft nicht zu [6] trifft zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 2 | 8 | 8 |
------------------------
================================================================================

================================================================================
Frage 27
--------------------------------------------------------------------------------
Im Gespräch mit SmartWeb fühlte ich mich ...
... unwohl [6] ... wohl
-------------------------
```

```
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 2 | 0 | 3 | 7 | 5 | 1 |
-------------------------
================================================================================

================================================================================
Frage 28
--------------------------------------------------------------------------------
Ich bin von der Leistung von SmartWeb
... enttï¿½uscht [6] ... beeindruckt
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 4 | 6 | 4 | 3 | 1 | 0 |
-------------------------
================================================================================

================================================================================
Frage 29
--------------------------------------------------------------------------------
Der Umgang mit SmartWeb hat ...
mich gelangweilt [6] mir Spaß gemacht
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 2 | 0 | 5 | 3 | 6 | 2 |
-------------------------
================================================================================
```

8.4.2 Formative Evaluation – Report 2

```
SmartWeb Evaluation

20.04.2007     Hannes Mögele, Florian Schiel

Version Prototyp:  0.6
Endgerät:          MDA Pro
Server:            2 x ( 2,8GHz, 3GB RAM, SuSE 9.2 )

Einstellung am SW-Client und experimentelles Set-up:
- Mikrophon: MDA Pro Headset und eingebautes Mikrophon (Freisprechen); z. T. mit
  wechselndem Abstand zum Mikrophon
- GPS-Lokalisation: Berlin und München
- Verbindungstyp: WLAN (44) und UMTS (6)
- Spracherkenneraktivierung: push-to-talk (Mikrophon muss für jede Anfrage
  aktiviert werden) und offene Spracherkennung
- Umgebung: outdoor und indoor
- Haltung des Endgerätes: zugeklappt und in hochformatiger Position
- Positionierung des Endgerätes: auf dem Tisch, in der Hand und wechselnd


Begleitende Evaluation - Report 2 (05.02.2007 - 18.04.2007)

Dieser Report berichtet über Mängel beim Betrieb des SmartWeb-Prototypen, die
während der begleitenden Evaluation in den bis zum Berichtszeitpunkt
durchgeführten Evaluationsexperimenten aufgetreten sind. Mängel-Reports der
begleitenden Evaluation sind mit 'Bii.x' (ii = Nummer der Kategorie,
x = fortlaufende Nummer der Subkategorie) bezeichnet.
Bitte beziehen Sie sich in der Korrespondenz auf diesen Bezeichner.


Als Quellen für diesen Report wurden
(1) die obligatorischen Bewertungen der Sessions durch die jeweiligen
    Versuchpersonen (vgl. hierzu den Abschnitt 'Fragebogen'),
(2) die freien Kommentare der Versuchspersonen (fakultativ),
(3) die Beobachtungen der Versuchsleitung (VL) und Aussagen der Versuchspersonen
    gegenüber der Versuchsleitung und
```

(4) die Ergebnisse der Spracherkennerbewertung
herangezogen.

(2) beinhaltet Anmerkungen der Versuchspersonen (im Bericht mit '"..."' markiert) , die diese zusätzlich zum Fragebogen für besonders erwähnenswert hielten; deshalb werden auch Anmerkungen berichtet, die nur in einer Session und von einer Versuchsperson geäußert wurden.

Aus (2) und (3) wurden bisher folgende Klassen mit Unterkatergorien gebildet (vgl. hierzu Report-1):

B01 Medientypen
B01.1 [Medientypen] fehlende Anzeige der gefundenen Medientypen

B02 Performanz
B02.1 [Performanz] zu langsame Verarbeitungsgeschwindigkeit
B02.2 [Performanz] ungenügende Bearbeitungsgenauigkeit

B03 Spracherkennung
B03.1 [Spracherkennung] ungenügende Spracherkennnerleistung
B03.2 [Spracherkennung] Out-Of-Vocabulary (neu)

B04 Sprachsynthese
B04.1 [Sprachsynthese] unflüssige Sprachsynthese
B04.2 [Sprachsynthese] schlecht verständliche Sprachsynthese
B04.3 [Sprachsynthese] fehlende Sprachsynthese
B04.4 [Sprachsynthese] Persona
B04.5 [Sprachsynthese] fehlerhafte Sprachsynthese (neu)

B05 GUI
B05.1 [GUI] nicht intuitiv
B05.2 [GUI] Anzeige des Bearbeitungsstatus
B05.3 [GUI] Bedienung
B05.4 [GUI] Mikrofonsymbol
B05.5 [GUI] Übersichtlichkeit

B06 Motivation / Emotion
B06.1 [Motivation/ Emotion] Frustration aufgrund von B02.2, B03.1

B07 Semantische Analyse
B07.1 [Semantische Analyse] verwirrende Ausgabe der semantischen Analyse

Für diesen Report wurden 50 Evaluationsexperimente mit 10 Versuchspersonen berücksichtigt. Davon waren 3 erste, 3 zweite, 7 dritte, 9 vierte, 8 fünfte, 8 sechste, 7 siebte und 5 achte Sitzungen.

Sprecher	Session	Aufnahmedatum
AAAD	e065	2007-04-04
AAAD	e067	2007-04-11
AAAD	e071	2007-04-18
AAAJ	e029	2007-02-09
AAAJ	e030	2007-02-12
AAAJ	e035	2007-02-23
AAAJ	e049	2007-03-20
AAAJ	e042	2007-03-07
AAAJ	e007	2007-02-05
AAAM	e008	2007-02-06
AAAM	e012	2007-02-13
AAAM	e011	2007-02-22
AAAM	e037	2007-02-27
AAAT	e028	2007-02-08
AAAT	e026	2007-02-14
AAAT	e034	2007-02-19
AAAT	e044	2007-03-09
AAAT	e052	2007-03-22

```
 AAAT      | e058   | 2007-03-27
-----------------------------------
 AAAW      | e010   | 2007-02-22
 AAAW      | e038   | 2007-02-28
 AAAW      | e040   | 2007-03-06
 AAAW      | e045   | 2007-03-12
 AAAW      | e061   | 2007-03-28
 AAAW      | e068   | 2007-04-16
-----------------------------------
 AADA      | e033   | 2007-02-19
 AADA      | e039   | 2007-02-28
 AADA      | e041   | 2007-03-07
 AADA      | e043   | 2007-03-09
 AADA      | e062   | 2007-03-28
-----------------------------------
 AADD      | e027   | 2007-02-08
 AADD      | e051   | 2007-03-21
 AADD      | e055   | 2007-03-26
 AADD      | e063   | 2007-03-29
 AADD      | e070   | 2007-04-18
-----------------------------------
 AAGW      | e047   | 2007-03-13
 AAGW      | e050   | 2007-03-20
 AAGW      | e057   | 2007-03-27
 AAGW      | e064   | 2007-04-03
 AAGW      | e066   | 2007-04-10
-----------------------------------
 ADMJ      | e013   | 2007-02-14
 ADMJ      | e046   | 2007-03-12
 ADMJ      | e048   | 2007-03-19
 ADMJ      | e054   | 2007-03-26
-----------------------------------
 AGJG      | e031   | 2007-02-15
 AGJG      | e032   | 2007-02-16
 AGJG      | e036   | 2007-02-23
 AGJG      | e053   | 2007-03-23
 AGJG      | e056   | 2007-03-27
 AGJG      | e069   | 2007-04-17
===================================


### POSITIVE ANMERKUNGEN DER VERSUCHSPERSONEN ################################

* Wettervorhersage korrekt
* Anzeigen der Karte auf die Anfrage: "Zeige mir Hotels in Hamburg"
* Feedback ist "viel besser als vorher"
* VP ist motiviert und der Umgang mit SmartWeb macht ihr Spaß
  "Es ist sehr interessant, mit so einem Gerät zu arbeiten"
* Fragen zum Kinoprogramm wurden erkannt und richtig beantwortet
* Beantwortung der Fragen zu Schauspielern aus verschiedenen Filmen
* Darstellung des Stadtplanausschnittes mit eingezeichneten Restaurantes und das
  Anzeigen der Restaurantesnamen ducrh Klicken auf die entsprechenden Symbole

============================================================================



### MÄNGEL ###################################################################

============================================================================
B01 Medientypen
============================================================================
B01.1 [Medientypen] fehlende Anzeige der gefundenen Medientypen
----------------------------------------------------------------------------
* gezielte Anfrage nach Web-Cam nicht beantwortet, trotz früherem
  erfolgreichem Versuch (AAAJ_e049)
* eine angezeigte Karte konnte nicht gezoomt werden (AAAT_e028)
* gefundene Texte können nicht immer angezeigt werden (AAAT_e028)
* bei der Frage nach Bildern vom Brandenburger Tor ist die
  Antwort: 'Webcambilder gefragt' und 'Resultat gefiltert' aber
  es kommt kein Bild (AADD_e027)
```

```
================================================================================
B02 Performanz
================================================================================
B02.1 [Performanz] zu langsame Verarbeitungsgeschwindigkeit
--------------------------------------------------------------------------------
* "langsame Bearbeitungsgeschwindigkeit" (AAAW_e010)
* "sehr lange Wartezeit auf Antwort" (AAAJ_e030)
* VP missfällt die starke Zeitverzögerung beim Aufbau von Karten (AADA_e062)
* beim langen Warten auf Ergebnisse (Route, Karte, Kinoprogramm) wird VP
  ungeduldig und gelangweilt (ADMJ_e054, AGJG_e036)

================================================================================
02.2 [Performanz] ungenügende Bearbeitungsgenauigkeit
--------------------------------------------------------------------------------
* "Trotz Korrektur hat das System keine Antwort geliefert." (AAAJ_e007)
* "schlechte Ergebnisausbeute" (AAAM_e008)
* "wenig Information und schlechte Auswahl an Internetseiten" (AAAM_e011)
* "Information und Sprachausgabe sehr mager" (AAAM_e011)
* "Smartweb ignoriert Ortsangaben" (AGJG_e036)
* Ausgabe Hauptstadt der Slowakei 'Wien' (AAAJ_e049)
* bei Zahlenangaben kommen immer völlig irrelevante Antworten, z.B. wird auf
  die Frage nach den aktuellen Kinopreisen u.a. eine Mobilfunk-Nummer angezeigt
  (AAAM_e011)
* die Frage 'gegen wen gewann Brasilien im Finale?' wurde mit 'Ballack'
  beantwortet (AAAM_e037)
* auf die Frage 'wo gibt es in München Kinos'? erscheint im Display 'uups'
  (AAAT_e026)
* auf die Frage nach dem Fernsehturm war die Ausgabe 'anzeige@TV-tower', und
  auf die Frage nach Museen 'anzeige@museum' (AAAT_e034)
* auf die Frage nach dem Naturkundemuseum kamen Antworten zu Naturkundemuseum
  in anderen Städten (AAAT_e034)
* Antwort auf die Frage nach dem erfolgreichsten Fußballspieler 2006 lieferte
  'Bill Gates' (AAAD_e067)
* die Kurzantwort ist oft unvollständig oder irreführend (auch wenn die Texte
  dazu die richtige Antwort liefern); z.B.: Frage: 'Wann war die Schweiz im
  Endspiel?', Antwort: 'um 11' (AGJG_e032)
* Auf die Frage nach Stadtführungen kamen Texte zu Berlin, aber nichts über
  Stadtführungen. Auf die Frage 'Wo kann ich eine Stadtrundfahrt machen?' war
  die Antwort 'Polen'. Auf die Frage 'Wo gibt es Cafes?' war die Antwort
  'Bildatlas Special San Francisco' (AGJG_e056)

================================================================================
B03.1 [Spracherkennung] ungenügende Spracheerkennnerleistung
--------------------------------------------------------------------------------
* selten sofortige, korrekte Spracherkennung (ADMJ_e046, AAAW_e045, AAGW_e050,
  AAAT_e034, AAAT_e052, AAAT_e058, AGJG_e056)
* ohne Headset ist die Spracherkennerleistung von SW schlecht (AAAT_e044)
* die Anfrage "zeige mir.." funktioniert besser als "wo ist.." (AADA_e062)
* "Spracherkennung ziemlich schlecht" (AAGW_e047)
* "Es fällt mir nichts mehr ein, was er verstehen könnte." (AAAW_e040)
* "versteht "Wendelin" statt "Berlin" (AADA_e033)
* Spracheingaben der VP wurden generell schlecht verstanden, das Mikrofonsymbol
  blieb grün. Wenn es auf orange wechselt, erfolgt trotzdem keine weitere
  Verarbeitung (AADD_e051, AADD_e070)
================================================================================
B03.2 [Spracherkennung] Out-Of-Vocabulary
--------------------------------------------------------------------------------
* es gibt in dieser Sitzung nur ein einziges Mal eine Sprachausgabe, die da
  lautete "OOV". Diese Ausgabe war unter anderem auch immer zu lesen, wenn
  statt einer Antwort die interne Suchtstrategie auf dem Display zu lesen war
  (AAAJ_e029)
* die Ausgabe "OOV" wurde verbal und auf dem Bildschirm ausgegeben (AAAM_e012)
* zu "OOV" werden 4/6 Texte angezeigt, die restlichen beiden Texte handeln von
  'Punks United' wobei die Frage "Gibt es in Berlin ein Ibis-Hotel?" war
  (AAAM_e012)
* die Ausgabe 'OOV' + 'Parser error' wured auf dem Display ausgegeben (AAAT_e034)

================================================================================
B04 Sprachsynthese
--------------------------------------------------------------------------------
```

```
================================================================================
B04.1 [Sprachsynthese] unflüssige Sprachsynthese
--------------------------------------------------------------------------------

================================================================================
B04.2 [Sprachsynthese] schlecht verständliche Sprachsynthese
--------------------------------------------------------------------------------
* Sprachaugabe unverständlich (hier: vermutlich wegen lauter Umgebung) (ADMJ_e046)

================================================================================
B04.3 [Sprachsynthese] fehlende Sprachsynthese
--------------------------------------------------------------------------------
* nur der Begrüßungstext wurde synthetisiert (AAAT_e044)

================================================================================
B04.3 [Sprachsynthese] Persona
--------------------------------------------------------------------------------
* "Die Sprachausgabe ist schrecklich!" (AAAW_e045)
* "Dieses Ding kann nicht sprechen!" (AAAW_e061
* "Stimme und Aussprache der englischen Wörter negativ." (ADMJ_e054)

================================================================================
B04.5 [Sprachsynthese] fehlerhafte Sprachsynthese
--------------------------------------------------------------------------------
* "Entschuldigung, wie bitte?" 3mal wiederholt (AAAW_e045)
* Sprachausgabe sehr selten und unvollständig (ADMJ_e046)

================================================================================
B05 GUI
================================================================================
B05.1 [GUI] nicht intuitiv
--------------------------------------------------------------------------------
* "Bei Einbgabeverbesserung mit Stift war mir das return zum Starten einer
  neuen Suche nicht intuitiv bewußt" (AAAJ_e029) (Anm.: trotz Instruktion)
* VP durch Anzeige "Rückfrage" verwirrt (AAAJ_e049)
* VP klickte anfangs immer wieder auf das Mikrofonsymbol, obwohl die Sitzung
  bei offenem Mikro stattfand. VP legte dann von sich aus den Stift weg
  (AAAT_e058) (Anm.: trotz Instruktion)
* "Das Auf- und Abscrollen innerhalb der Antworten sollte erleichtert werden"
  (AADA_e033)
* Wenn SmartWeb etwas richtig gefunden hat, dann sollte man dazu einen Verweis
  für nähere Informationen erhalten (AADA_e041)
* Obwohl das Symbol für Off-Talk inzwischen bekannt ist, ist VP verwirrt, weil
  es zu völlig unpassenden Gelegenheiten erscheint (AADD_e055)

================================================================================
B05.2 [GUI] Anzeige des Bearbeitungsstatus
--------------------------------------------------------------------------------
* Bearbeitungszustandsanzeige für VP nicht ausreichend oder sogar verwirrend
  bis entnervend (AAAJ_e007, AAAJ_e029, AAAJ_e035, AAAM_e008, AAAT_e052,
  AAAT_e058, AADD_e055, AADD_e070, AAGW_e066, ADMJ_e054)
* "Es ist nervig, dass man nicht sieht, hat er das registriert oder nicht"
  (AAAJ_e030)
* "es ist nicht ersichtlich, ob er was macht oder nicht, da der blaue Punkt
  nicht immer zu sehen ist und manchmal 'keine Antwort' oder 'noch keine
  Antwort' ausgegeben wird, manchmal auch nicht." (AAAJ_e042)
* "Es ist absolut nicht erkennbar ob und wann was passiert." (AAAJ_e049)
* "Die kleine blaue Kugel am oberen Bildschirmrand war nicht da. Auch sonst war
  nicht erkennbar, ob die Anfrage noch weiter bearbeitet wird" (AGJG_e069)

================================================================================
B05.3 [GUI] Bedienung
--------------------------------------------------------------------------------
* VP bemängelt fehlenden Suchstatus (AAAJ_e035)
* "Korrektur mit Stift wurde einfach vom System unterbrochen, wenn dieses ein
  Ergebnis oder eine Fehlermeldung geliefert hat" (AAAJ_e049)
* VP bedauert die nicht gegebene Möglichkeit, auf die ausgegebenen Antworten zu
  klicken und dazu weitere Informationen zu bekommen (AAAJ_e007)
* "Ich kann die Recherche nicht selbst abbrechen" (AAAJ_e007)
* es war heute nicht möglich auf dem Display nach unten zu scrollen (AAAM_e008)
```

```
* VP durch unlogisch erscheinendes Off-Talk Symbol amüsiert (AAAT_e052)
* "es gibt keine Funktion 'zurück', bei der man einfach wieder auf die
  vorherige Antwort springen kann" (AADA_e062)
* "Wie kann ich die Suche stoppen, wenn System zu lange sucht?" (AAAW_e068)
* "Negativ ist die steigende Unzuverlässigkeit bzw. lange Wartezeit"
  (AAAW_e068)
* "Warum gibt es keinen 'Stop-Button', um die Anfrage zu beenden?" (AAGW_e057)

================================================================================
B05.4 [GUI] Mikrofonsymbol
--------------------------------------------------------------------------------
* das Mikrofonsymbol bleibt oft grün, obwohl die VP deutlich und nah am Gerät
  spricht (AAAM_e011, AADA_e041, AAAJ_e035, AADD_e051, AGJG_e031)
* manchmal erscheint blaues Mikrofonsymbol, nach dem erkannten Input wurde
  Mikrofonsymbol wieder grün, VP wusste nicht, ob Deaktivieren erforderlich
  oder weitere Anfrage möglich ist (ADMJ_e046, AADA_e041, AADD_e027)
* offenes MS für VP neu, daher deaktiviert die VP anfangs oft grünes
  Mikrofonsymbol und aktiviert es wieder (AAAW_e045) (Anm.: trotz Instruktion)
* das Off-Talk-Symbol war oft zu sehen, aber ohne erkennbare Logik - es
  erscheint auch, wenn es still ist oder gerade keine Spracherkennung läuft
  (AAAJ_e035)
* "lieber drücke ich mit dem Stift auf das MS" (AAAJ_e049)
* VP stellt oft Anfragen, wenn das Mikro noch nicht aktiviert ist (AAAD_e067)
* "Es ist unklar, wann das Mikrofonsymbol aktiv ist." (ADMJ_e013)

================================================================================
B05.5 [GUI] Übersichtlichkeit
--------------------------------------------------------------------------------
* "Infos auf dem Display recht unübersichtlich, angezeigte Karte war nicht
  anklickbar." (AAAM_e008)
* verwirrende Ausgabe: 'keine Antwort' + 'Karte gefragt' + 'Recherche
  abgeschlossen' (AAAJ_e029)

================================================================================
B06 Motivation / Emotion
================================================================================
B06.1 [Motivation/ Emotion] Frustration aufgrund von B02.2, B03.1
--------------------------------------------------------------------------------
* VP wird immer frustrierter: "es ist alles so unkontrolliert" (AAAM_e008)
* VP frustriert/ gelangweilt, nachdem keine passenden Ergebnisse kommen
  (AAAT_e026, AAAT_e034, AAAT_e044, AAAT_e052, AAAT_e058, AADA_e039, AADD_e070,
  AADD_e070, AGJG_e069)
* "Wie kann ich das jetzt am besten formulieren, dass der mich versteht."
  (AAAT_e052)
* "Ich hab das Gefühl, das System wird von Mal zu Mal schlechter." (AAAJ_e029)
* "das macht keinen Spaß heute" (AADA_e039)
* "Frustierend ist das ständige Wiederholen der Fragen." (AAAD_e067)
* "Es ist ein bißchen frustierend, wenn er nicht versteht." (AAAD_e071)

================================================================================
B07 Semantische Analyse
================================================================================
B07.1 [Semantische Analyse] verwirrende Ausgabe der semantischen Analyse
--------------------------------------------------------------------------------



### FRAGEBOGEN ###############################################################

Es wird jeweils die laufende Nummer der Fragen (Frage 1), das Item
('Die Bedienung von SmartWeb ist ...') mit Attributen ('schwierig' - 'leicht')
gegeben. Die Ziffer zwischen den Attributen gibt die Skalierung ('[5]' oder
'[6]') wieder.

Für die Auswertung der Fragebögen wurden alle bisherigen Experimente
berücksichtigt; also auch die, über die bereits im Report-1 'Begleitende_1.txt'
berichtet wurde.  Für die Session 'e028' (AAAT) liegt keine Fragebogen vor.

================================================================================
Frage 1
--------------------------------------------------------------------------------
```

```
Die Bedienung von SmartWeb ist ...
schwierig [6] leicht
---------------------------
| 1 | 2 | 3 | 4  | 5  | 6 |
|---|---|---|----|----|---|
| 1 | 5 | 9 | 20 | 24 | 8 |
---------------------------
================================================================================

================================================================================
Frage 2
--------------------------------------------------------------------------------
Die Lesbarkeit der Schrift war...
sehr schlecht [6] sehr gut
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 0 | 0 | 2 | 5 |35 |25 |
-------------------------
================================================================================

================================================================================
Frage 3
--------------------------------------------------------------------------------
Die Hervorhebungen erleichterten die Bedienung.
stimme ich nicht zu [6] stimme ich zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 3 |16 |11 |16 |12 | 0 |
-------------------------
keine Bewertung: 9
================================================================================

================================================================================
Frage 4
--------------------------------------------------------------------------------
Die Anordnung der Informationen auf dem Display finde ich...
verwirrend [6] übersichtlich
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 2 | 7 |17 |17 |20 | 4 |
-------------------------
================================================================================

================================================================================
Frage 5
--------------------------------------------------------------------------------
Die Ausdrucksweise von SmartWeb war...
inkonsistent [6] konsistent
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 6 |14 | 8 |21 |12 | 5 |
-------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 6
--------------------------------------------------------------------------------
SmartWeb zeigt mir, wenn es beschäftigt ist.
nie [6] immer
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 2 |16 |11 |18 |12 | 6 |
-------------------------
keine Bewertung: 2
================================================================================
```

```
================================================================================
Frage 7
--------------------------------------------------------------------------------
Die Fehlermeldungen von SmartWeb sind...
wenig hilfreich [6] sehr hilfreich
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
|16 | 25| 7 | 5 | 6 | 0 |
-------------------------
keine Bewertung: 8
================================================================================

================================================================================
Frage 8
--------------------------------------------------------------------------------
Die Geschwindigkeit von SmartWeb fand ich...
zu langsam [5] genau richtig
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
|30 |26 | 9 | 2 | 0 |
--------------------
================================================================================

================================================================================
Frage 9
--------------------------------------------------------------------------------
Das SmartWeb funktionierte...
unzuverlässig [6] reibungslos
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
|22 |24 | 8 |10 | 3 | 0 |
-------------------------
================================================================================

================================================================================
Frage 10
--------------------------------------------------------------------------------
SmartWeb lieferte...
zu viele Informationen [6] zu wenig Informationen
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 0 | 1 | 5 | 9 |21 |26 |
-------------------------
keine Bewertung: 5
================================================================================

================================================================================
Frage 11
--------------------------------------------------------------------------------
So wie SmartWeb in der heutigen Sitzung funktioniert hat, könnte auch jeder
andere damit umgehen.
stimme ich nicht zu [6] stimme ich zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
|32 |19 | 4 | 3 | 7 | 2 |
-------------------------
================================================================================

================================================================================
Frage 12
--------------------------------------------------------------------------------
Fehleingaben konnte ich leicht korrigieren.
stimme ich nicht zu [6] stimme ich zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
```

```
|---|---|---|---|---|---|
| 0 |10 | 7 |13 |28 | 8 |
-------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 13
--------------------------------------------------------------------------------
Wie angenehm fanden Sie die Stimme?
sehr unangenehm [5] sehr angenehm
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 4 |27 |18 |12 | 1 |
--------------------
keine Bewertung: 5
================================================================================

================================================================================
Frage 14
--------------------------------------------------------------------------------
Welche Anstrengung war nötig, um die Äußerungen zu verstehen?
selbst größte Anstrengung reicht nicht zum Verstehen [5] es war keine
Anstrengung zum Verstehen erforderlich
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 1 | 8 |19 |24 |10 |
--------------------
keine Bewertung: 5
================================================================================

================================================================================
Frage 15
--------------------------------------------------------------------------------
Wie würden Sie die Natürlichkeit der Stimme einschätzen?
sehr unnatürlich [5] sehr natürlich
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
|18 |16 |13 |14 | 0 |
--------------------
keine Bewertung: 6
================================================================================

================================================================================
Frage 16
--------------------------------------------------------------------------------
Fanden Sie heute bestimmte Wörter schwer zu verstehen?
ständig [5] nie
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 2 | 7 |21 |14 |15 |
--------------------
keine Bewertung: 8
================================================================================

================================================================================
Frage 17
--------------------------------------------------------------------------------
Wie würden Sie insgesamt die Sprachqualität der gehörten Äußerungen beurteilen?
sehr schlecht [5] sehr gut
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| 7 |22 |20 |12 | 1 |
--------------------
keine Bewertung: 5
================================================================================
```

```
================================================================================
Frage 18
--------------------------------------------------------------------------------
SmartWeb konnte meine Fragen beantworten.
keine einzige [6] praktisch alle
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
|21 |25 | 8 | 3 | 9 | 0 |
-------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 19
--------------------------------------------------------------------------------
SmartWeb hat mich schnell zur gewï¿½nschten Information gefï¿½hrt.
trifft nicht zu [6] trifft zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
|30 |25 | 5 | 2 | 5 | 0 |
-------------------------
================================================================================

================================================================================
Frage 20
--------------------------------------------------------------------------------
SmartWeb hat mir die richtigen Informationen geliefert.
trifft nicht zu [6] trifft zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
|19 |22 | 9 |12 | 1 | 2 |
-------------------------
keine Bewertung: 2
================================================================================

================================================================================
Frage 21
--------------------------------------------------------------------------------
Der Gesprächsverlauf wurde eher von...
SmartWeb bestimmt [6] mir selber bestimmt
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 7 |10 | 3 | 5 |10 |32 |
-------------------------
================================================================================

================================================================================
Frage 22
--------------------------------------------------------------------------------
Ich wusste immer, wie ich die nächste Eingabe (per Sprache, Tastatur, Stift)
machen konnte.
trifft nicht zu [6] trifft zu
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 4 | 4 | 5 | 8 |31 |15 |
-------------------------
================================================================================

================================================================================
Frage 23
--------------------------------------------------------------------------------
Ich musste meine Eingaben wiederholen.
praktisch jedes mal [6] nie
-------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
```

```
|---|---|---|---|---|---|
|27 |25 | 8 | 3 | 3 | 0 |
-------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 24
--------------------------------------------------------------------------------
Die Pausen zwischen Eingabe und Antwort erschienen mir ...
... sehr lang [5] ... sehr kurz
--------------------
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
|24 |23 |18 | 2 | 0 |
--------------------
================================================================================

================================================================================
Frage 25
--------------------------------------------------------------------------------
Mit Hilfe der Spracheingabe komme ich schneller ans Ziel
trifft nicht zu [6] trifft zu
------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
|23 |17 |11 | 8 | 6 | 1 |
------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 26
--------------------------------------------------------------------------------
Die Kombination von Stift- und Spracheingabe finde ich sinnvoll.
trifft nicht zu [6] trifft zu
------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 0 | 0 | 4 | 8 |19 |36 |
------------------------
================================================================================

================================================================================
Frage 27
--------------------------------------------------------------------------------
Im Gespräch mit SmartWeb fühlte ich mich ...
... unwohl [6] ... wohl
------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 5 | 5 |13 |21 |18 | 5 |
------------------------
================================================================================

================================================================================
Frage 28
--------------------------------------------------------------------------------
Ich bin von der Leistung von SmartWeb
... enttï¿½uscht [6] ... beeindruckt
------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
|19 |22 |13 | 8 | 4 | 1 |
------------------------
================================================================================

================================================================================
Frage 29
--------------------------------------------------------------------------------
Der Umgang mit SmartWeb hat ...
```

```
mich gelangweilt [6] mir Spaß gemacht
------------------------
| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 8 | 5 |17 |14 |14 | 9 |
------------------------
================================================================================


### SPRACHERKENNEREVALUATION ###################################################

Die Auswertung der Spracherkennerergebnisse liefert eine Correctness von 45.28%
und eine Accuracy von 39.55% bei N=6720 (H=3043, D=2031, S=1646, I=385).

================================================================================
```

8.4.3 Formative Evaluation – Report 3

```
SmartWeb Evaluation

01.08.2007      Hannes Mögele, Florian Schiel

Version Prototyp:  1.0
Endgerät:          MDA Pro
Server:            2 x ( 2,8GHz, 3GB RAM, SuSE 10.1 )

Einstellung am SW-Client und experimentelles Set-up:
- Mikrophon: MDA Pro Headset und eingebautes Mikrophon (Freisprechen); z. T. mit
  wechselndem Abstand zum Mikrophon
- GPS-Lokalisation: Berlin und München
- Verbindungstyp: WLAN (11) und UMTS (11)
- Spracherkenneraktivierung: push-to-talk (Mikrophon muss für jede Anfrage
  aktiviert werden) und offene Spracherkennung
- Umgebung: outdoor und indoor
- Haltung des Endgerätes: zugeklappt und in hochformatiger Position
- Positionierung des Enderätes: auf dem Tisch, in der Hand und wechselnd


Begleitende Evaluation - Report 3 (19.04.2007 - 18.07.2007)

Dieser Report berichtet über Mängel beim Betrieb des SmartWeb-Prototypen, die
während der begleitenden Evaluation in den bis zum Berichtszeitpunkt
durchgeführten Evaluationsexperimenten aufgetreten sind. Mängel-Reports der
begleitenden Evaluation sind mit 'Bii.x' (ii = Nummer der Kategorie,
x = fortlaufende Nummer der Subkategorie) bezeichnet.
Bitte beziehen Sie sich in der Korrespondenz auf diesen Bezeichner.


Als Quellen für diesen Report wurden
(1) die obligatorischen Bewertungen der Sessions durch die jeweiligen
    Versuchpersonen (vgl. hierzu den Abschnitt 'Fragebogen'),
(2) die freien Kommentare der Versuchspersonen (fakultativ) und
(3) die Beobachtungen der Versuchsleitung (VL) und Aussagen der Versuchspersonen
    gegenüber der Versuchsleitung
herangezogen.

(2) beinhaltet Anmerkungen der Versuchspersonen (im Bericht mit '"..."'
markiert) , die diese zusätzlich zum Fragebogen für besonders erwähnenswert
hielten; deshalb werden auch Anmerkungen berichtet, die nur in einer Session
und von einer Versuchsperson geäußert wurden.

Aus (2) und (3) wurden bisher folgende Klassen mit Unterkatergorien gebildet
(vgl. hierzu Report-1, Report-2):

B01 Medientypen
B01.1 [Medientypen] fehlende Anzeige der gefundenen Medientypen
```

B02 Performanz
B02.1 [Performanz] zu langsame Verarbeitungsgeschwindigkeit
B02.2 [Performanz] ungenügende Bearbeitungsgenauigkeit

B03 Spracherkennung
B03.1 [Spracherkennung] ungenügende Spracherkennnerleistung
B03.2 [Spracherkennung] Out-Of-Vocabulary

B04 Sprachsynthese
B04.1 [Sprachsynthese] unflüssige Sprachsynthese
B04.2 [Sprachsynthese] schlecht verständliche Sprachsynthese
B04.3 [Sprachsynthese] fehlende Sprachsynthese
B04.4 [Sprachsynthese] Persona
B04.5 [Sprachsynthese] fehlerhafte Sprachsynthese

B05 GUI
B05.1 [GUI] nicht intuitiv
B05.2 [GUI] Anzeige des Bearbeitungsstatus
B05.3 [GUI] Bedienung
B05.4 [GUI] Mikrofonsymbol
B05.5 [GUI] Übersichtlichkeit

B06 Motivation / Emotion
B06.1 [Motivation/ Emotion] Frustration aufgrund von B02.2, B03.1

B07 Semantische Analyse
B07.1 [Semantische Analyse] verwirrende Ausgabe der semantischen Analyse

Für diesen Report wurden 22 Sessions von 8 Versuchspersonen herangezogen.
Dabei waren 1 vierte, 2 fünfte, 2 sechste, 2 siebte, 2 achte, 8 neunte und 5 zehnte Sitzungen

s_spkcode	s_session_name	s_recordingdate
AAAD	e074	2007-04-25
AAAD	e075	2007-05-02
AAAD	e076	2007-05-09
AAAD	e078	2007-05-16
AAAD	e081	2007-05-23
AAAD	e082	2007-06-06
AAAD	e085	2007-06-20
AGJG	e072	2007-04-19
AGJG	e073	2007-04-23
AGJG	e086	2007-06-25
AGJG	e096	2007-07-18
AADA	e091	2007-07-02
ADMJ	e077	2007-05-14
ADMJ	e080	2007-05-23
ADMJ	e089	2007-06-27
ADMJ	e089	2007-06-28
AAAW	e079	2007-05-21
AADD	e090	2007-06-28
AAAT	e092	2007-07-03
AAAT	e093	2007-07-09
AAAJ	e094	2007-07-10
AAAJ	e095	2007-07-13

POSITIVE ANMERKUNGEN DER VERSUCHSPERSONEN

```
* Prototyp 0.8 erkennt die Anfragen wesentlich besser als 0.6 (AAAD_e075)
* Sprachausgabe bei 0.8 deutlich besser als bei 0.6 (AAAD_e076)
* Spracherkennung bei UMTS-Aufnahme wesentlich besser als bei WLAN-Aufnahme,
  trotz lauter Umgebung (AAAD_e078)
* "Wow, bin beeindruckt!" VP überrascht von guten Ergebnissen. Auf Anfrage nach
  italienischem Restaurant in Duisburg liefert SmartWeb Karte von Duisburg mit
  eingetragenen italienischen Restaurants (AAAD_e081)
* beim Anklicken des Icones des Arztes sagte System den Namen des Arztes
  (AAAD_e082)
* VP motiviert, mit Spaß bei der Sache, experimentiert (AADA_e091; AAAJ_e095)
* Die Karten mit Ärzten ist zoombar und es können Wege (bzw. Linien)
  eingezeichnet werden, was die VP eigenständig herausfand (AADA_e091)

================================================================================



#### MÄNGEL ##################################################################

================================================================================
B01 Medientypen
================================================================================
B01.1 [Medientypen] fehlende Anzeige der gefundenen Medientypen
--------------------------------------------------------------------------------

================================================================================
B02 Performanz
================================================================================
B02.1 [Performanz] zu langsame Verarbeitungsgeschwindigkeit
--------------------------------------------------------------------------------

================================================================================
02.2 [Performanz] ungenügende Bearbeitungsgenauigkeit
--------------------------------------------------------------------------------
* von zwölf Fragen wurden lediglich 4 beantwortet und diese alle falsch
  (AGJG_e074)
* System versteht keine Pluralformen (AAAW_e079)
* es wurden sehr wenige Informationen ausgegeben und wenn, dann auch falsche
  oder irrelevante (AGJG_e086)
* Auf die Frage, wo es Ärzte gibt, erscheint eine Karte mit Symbolen, die
  angeklickt werden können. Es werden Namen angezeigt, allerdings keine
  Fachrichtung, Titel, Vorname etc. (AGJG_e086)
* es wurde nur eine einzige Frage richtig beantwortet. Andere Antworten (wenn
  überhaupt welche gegeben wurden), hatten nichts mit der Frage zu tun
  (AADD_e090)
* es wurde nur eine Frage beantwortet und diese nicht sehr ausführlich
  (AAAJ_e094)
* die meisten Fragen - außer einer - überhaupt nicht beantwortet (AGJG_e096)
* bei Frage7 bleibt das Antwortfeld leer, aber es werden 6 Texte angezeigt, von
  denen einer richtig ist (AGJG_e096)

================================================================================
B03.1 [Spracherkennung] ungenügende Spracherkennnerleistung
--------------------------------------------------------------------------------
* es wurde fast keine Anfrage komplett richtig erkannt(AAAD_e076)
* schlechte Spracherkennung (AGJG_e072)
* SmartWeb erkannte einige Anfragen (teilweise) korrekt, antwortet aber nicht
  (AGJG_e074)
* System erkannte Weinberg statt Bamberg (AAAD_e078)
* VP redet sehr leise, wird oft vom System nicht verstanden (ADMJ_e081)
* alle Fragen wurden nicht oder falsch verstanden (AADD_e090)
* keine Frage wurde vollständig erkannt (AADA_e091)
* ungenügende Spracherkennung bis auf das Wort "Allgemeinmediziner", das auf
  Anhieb richtig erkannt wurde (AAAT_e092)
* keine Frage wird beim ersten Mal richtig erkannt (AAAT_e093; AGJG_e096)
* praktisch jede Frage musste korrigiert werden (AAAJ_e094)

================================================================================
B03.2 [Spracherkennung] Out-Of-Vocabulary
--------------------------------------------------------------------------------
```

```
================================================================================
B04 Sprachsynthese
================================================================================
B04.1 [Sprachsynthese] unflüssige Sprachsynthese
--------------------------------------------------------------------------------

================================================================================
B04.2 [Sprachsynthese] schlecht verständliche Sprachsynthese
--------------------------------------------------------------------------------
* "welche Sprache spricht der?" (AAAT_e092)

================================================================================
B04.3 [Sprachsynthese] fehlende Sprachsynthese
--------------------------------------------------------------------------------
* Sprachausgabe erfolgt nur ein einziges Mal bei Anfrage nach dem Wetter
  (AGJG_e073)

================================================================================
B04.3 [Sprachsynthese] Persona
--------------------------------------------------------------------------------

================================================================================
B04.5 [Sprachsynthese] fehlerhafte Sprachsynthese
--------------------------------------------------------------------------------
* "anscheinend zählt der irgendwelche Koordinaten auf." (AAAT_e092)
* Sprachausgabe scheint defekt zu sein - System sagt öfter mal "stern" in
  englischer Aussprache (AAAT_e093)

================================================================================
B05 GUI
================================================================================
B05.1 [GUI] nicht intuitiv
--------------------------------------------------------------------------------

================================================================================
B05.2 [GUI] Anzeige des Bearbeitungsstatus
--------------------------------------------------------------------------------
* "manchmal ist nicht klar, ob doch noch eine Antwort kommt" (AGJG_e072)
* SmartWeb zeigte oftmals gar keine Reaktion, es ist aber nicht klar, wie lange
  man auf eine Antwort warten sollte. Manchmal erscheint sehr spät doch noch
  eine Antwort (AGJG_e074)
* manchmal war System mit der Suche nach einer Antwort beschäftigt und
  reagierte nicht gleich auf nächste Anfrage, das war für VP verwirrend
  (AAAD_e078)
* "kommt noch was oder soll ich die Frage gleich noch mal stellen?"
  (AADD_eo090)
* VP deutete ständig auf den Touchscreen, die Reaktionen des Gerätes waren
  sehr verwirrend (AAAT_e092)
* es ist nie klar, ob man auf eine Antwort warten soll oder ob nichts mehr
  kommt (AAAT_e093; AAAJ_e095)
* "Man sollte die Frage nicht einmal bestätigen können, bevor SW mit der Suche
  beginnt." (AAAJ_e094)
* VP weiß nicht, wann SW wirklich beschäftigt ist und wann nichts mehr zu
  erwarten ist (AAAJ_e094)

================================================================================
B05.3 [GUI] Bedienung
--------------------------------------------------------------------------------
* es dauert sehr lange, bis ein Ergebnis ausgegeben wird (bei Kinovorstellung
  oder Routen-Anfrage) (AAAD_e076)
* auf zwei Drittel der Anfragen reagierte SmartWeb überhaupt nicht, generell
  dauert die Bearbeitung sehr lange. Es ist nicht nachvolziehbar, warum
  SmartWeb gerade auf bestimmte Webseiten zurückgreift (Bsp.
  www.zuckerarsch.org lieferte völlig belanglose Infos). Anfragen werden nur
  für das aktuelle Datum bearbeitet (AGJG_e073)
* "es dauert lange bis die Tastatur kommt"
* VP stellte immer sehr ähnliche Fragen und wurde mal verstanden und mal nicht
  (AGJG_e086)
* die VP wird generell vom System sehr schlecht verstanden. Das hat sich seit
  der ersten Sitzung auch nicht gebessert. (AADD_e090)
* BST wurde nicht angezeigt, daraufhin hat die VP das Gerät im Querformat mit
```

```
  normaler Tastatur benutzt. Die Karten mit Ärzten ist zoombar und es können
  Wege (bzw. Linien) eingezeichnet werden, was die VP eigenständig herausfand
  (AADA_e091)
* Das Gerät zeigte neues unverständliches Verhalten. In die Landkarte wurde
  gezoomt, obwohl das durch die VP nicht gewollt war. Sprachausgabe unbrauchbar
  (AAAT_e092)
* zur ersten Frage gibt es angeblich 59 Texte, aber es kann keiner angezeigt
  werden - Gerät wird mit der Antwortsuche nicht fertig (AAAT_e093)
* VP beklagt sich über die Langsamkeit von SW (AAAJ_e094)
* System arbeitet sehr langsam (AGJG_e096)

================================================================================
B05.4 [GUI] Mikrofonsymbol
--------------------------------------------------------------------------------
* blaues Mikrofonsymbol erschien sehr oft auch bei direkter Anfrage an das
  System und ohne Off-Talk, wirkte dadurch verwirrend (AAAD_e074)
* blaues Mikrofonsymbol tritt auf, nachdem das rote Mikro wieder
  durchgestrichen ist und die gestellte Anfrage wird nicht erkannt (no input)
  (AAAD_e075)
* blaues Mikrofonsymbol erscheint sehr häufig (AAAD_e076)
* ab und zu trat blaues Mikrofonsymbol auf (AAAD_e078)
* es dauerte zu lange, bis Mikrofonsymbol nach dem Anklicken aktiviert wurde
  (AAAD_e078)
* "vor allem das blaue Mikrofonsymbol" (AAAW_e079)
* bei UMTS Aufnahmen dauert es zu lange bis Mikrofonsymbol nach dem Anklicken
  aktiviert ist (ADMJ_e081)
* blaues Mikrofonsymbol erscheint bei ganz normalen Anfragen und diese werden
  als Off-talk verstanden und gar nicht bearbeitet (AAAD_e082)
* bei UMTS dauer es ein wenig länger bis grünes Mikrofonsymbol erscheint
  (AAAD_e085)
* laut VP vergeht mehr Zeit als in früheren Versuchen, bis das Mikrofonsymbol
  auf grün schaltet (AAAJ_e094)

================================================================================
B05.5 [GUI] Übersichtlichkeit
--------------------------------------------------------------------------------

================================================================================
B06 Motivation / Emotion
================================================================================
B06.1 [Motivation/ Emotion] Frustration aufgrund von B02.2, B03.1
--------------------------------------------------------------------------------
* Enttäuschung wegen vielfachen Absturzes, Frustration bei schlecht erkannten
  Anfragen (AAAD_e074)
* VP reagiert gefrustet auf nicht verstandene Anfragen
* Enttäuschung, da nur wenige Antworten korrekt waren
* Unzufriedenheit mit heutigen Ergebnissen (AAAD_e082)
* VP gelangweilt bis genervt (AAAT_e092)
* VP unmotiviert und frustriert (AAAT_e093)
* VP ist nach zehn Sitzungen extrem genervt von der schlechten Qualität des
  Systems. VP stellt nicht die Fragen, die sie interessieren, sondern solche
  bei denen eine richtige Antwort wahrscheinlich ist (AAAT_e093)
* VP wurde vom System sehr schlecht verstanden, aber aufgrund von Routine hat
  sie ohne Aufregung (nicht schneller, lauter, ärgerlicher etc.) ihre Anfragen
  gestellt, dennoch leicht resigniert (AGJG_e096)

================================================================================
B07 Semantische Analyse
================================================================================
B07.1 [Semantische Analyse] verwirrende Ausgabe der semantischen Analyse
--------------------------------------------------------------------------------

================================================================================




### FRAGEBOGEN ###################################################################

Es wird jeweils die laufende Nummer der Fragen (Frage 1), das Item
('Die Bedienung von SmartWeb ist ...') mit Attributen ('schwierig' - 'leicht')
```

gegeben. Die Ziffer zwischen den Attributen gibt die Skalierung ('[5]' oder '[6]') wieder.

Für die Auswertung der Fragebögen wurden alle bisherigen Experimente (88) berücksichtigt; also auch die, über die bereits im Report-1 und Report-2 berichtet wurde.

```
================================================================================
Frage 1
--------------------------------------------------------------------------------
Die Bedienung von SmartWeb ist ...
schwierig [6] leicht
-----------------------------
| 1 | 2 | 3  | 4  | 5  | 6  |
|---|---|----|----|----|----|
| 1 | 6 | 13 | 21 | 32 | 15 |
-----------------------------
keine Bewertung: 0
================================================================================

================================================================================
Frage 2
--------------------------------------------------------------------------------
Die Lesbarkeit der Schrift war...
sehr schlecht [6] sehr gut
---------------------------
| 1 | 2 | 3 | 4 | 5  | 6  |
|---|---|---|---|----|----|
| 0 | 0 | 4 | 8 | 45 | 31 |
---------------------------
keine Bewertung: 0
================================================================================

================================================================================
Frage 3
--------------------------------------------------------------------------------
Die Hervorhebungen erleichterten die Bedienung.
stimme ich nicht zu [6] stimme ich zu
-----------------------------
| 1 | 2  | 3  | 4  | 5  | 6 |
|---|----|----|----|----|---|
| 5 | 19 | 13 | 21 | 16 | 0 |
-----------------------------
keine Bewertung: 14
================================================================================

================================================================================
Frage 4
--------------------------------------------------------------------------------
Die Anordnung der Informationen auf dem Display finde ich...
verwirrend [6] übersichtlich
-----------------------------
| 1 | 2 | 3  | 4  | 5  | 6 |
|---|---|----|----|----|---|
| 3 | 7 | 18 | 28 | 28 | 4 |
-----------------------------
keine Bewertung: 0
================================================================================

================================================================================
Frage 5
--------------------------------------------------------------------------------
Die Ausdrucksweise von SmartWeb war...
inkonsistent [6] konsistent
-----------------------------
| 1 | 2  | 3  | 4  | 5  | 6 |
|---|----|----|----|----|---|
| 7 | 19 | 10 | 28 | 15 | 8 |
-----------------------------
keine Bewertung: 1
================================================================================
```

```
================================================================================
Frage 6
--------------------------------------------------------------------------------
SmartWeb zeigt mir, wenn es beschäftigt ist.
nie [6] immer
-----------------------------
| 1 | 2  | 3  | 4  | 5  | 6 |
|---|----|----|----|----|---|
| 2 | 20 | 16 | 21 | 20 | 7 |
-----------------------------
keine Bewertung: 2
================================================================================

================================================================================
Frage 7
--------------------------------------------------------------------------------
Die Fehlermeldungen von SmartWeb sind...
wenig hilfreich [6] sehr hilfreich
----------------------------
| 1  | 2  | 3  | 4 | 5 | 6 |
|----|----|----|---|---|---|
| 18 | 29 | 10 | 7 | 9 | 0 |
----------------------------
keine Bewertung: 15
================================================================================

================================================================================
Frage 8
--------------------------------------------------------------------------------
Die Geschwindigkeit von SmartWeb fand ich...
zu langsam [5] genau richtig
----------------------------
| 1  | 2  | 3  | 4 | 5 | 6 |
|----|----|----|---|---|---|
| 40 | 32 | 14 | 2 | 0 | 0 |
----------------------------
keine Bewertung: 0
================================================================================

================================================================================
Frage 9
--------------------------------------------------------------------------------
Das SmartWeb funktionierte...
unzuverlässig [6] reibungslos
-----------------------------
| 1  | 2  | 3  | 4  | 5 | 6 |
|----|----|----|----|---|---|
| 26 | 34 | 14 | 11 | 3 | 0 |
-----------------------------
keine Bewertung: 0
================================================================================

================================================================================
Frage 10
--------------------------------------------------------------------------------
SmartWeb lieferte...
zu viele Informationen [6] zu wenig Informationen
----------------------------
| 1 | 2 | 3 | 4  | 5  | 6  |
|---|---|---|----|----|----|
| 0 | 1 | 7 | 16 | 27 | 31 |
----------------------------
keine Bewertung: 6
================================================================================

================================================================================
Frage 11
--------------------------------------------------------------------------------
So wie SmartWeb in der heutigen Sitzung funktioniert hat, könnte auch jeder
andere damit umgehen.
```

stimme ich nicht zu [6] stimme ich zu

1	2	3	4	5	6
35	27	7	5	11	3

keine Bewertung: 0

Frage 12

Fehleingaben konnte ich leicht korrigieren.
stimme ich nicht zu [6] stimme ich zu

1	2	3	4	5	6
2	13	8	15	37	12

keine Bewertung: 1

Frage 13

Wie angenehm fanden Sie die Stimme?
sehr unangenehm [5] sehr angenehm

1	2	3	4	5
9	29	23	21	1

keine Bewertung: 5

Frage 14

Welche Anstrengung war nötig, um die Äußerungen zu verstehen?
selbst größte Anstrengung reicht nicht zum Verstehen [5] es war keine Anstrengung zum Verstehen erforderlich

1	2	3	4	5
3	14	25	31	10

keine Bewertung: 5

Frage 15

Wie würden Sie die Natürlichkeit der Stimme einschätzen?
sehr unnatürlich [5] sehr natürlich

1	2	3	4	5
24	22	20	16	0

keine Bewertung: 6

Frage 16

Fanden Sie heute bestimmte Wörter schwer zu verstehen?
ständig [5] nie

1	2	3	4	5
5	11	26	22	16

```
------------------------
keine Bewertung: 8
================================================================================

================================================================================
Frage 17
--------------------------------------------------------------------------------
Wie würden Sie insgesamt die Sprachqualität der gehörten Äußerungen beurteilen?
sehr schlecht [5] sehr gut
-------------------------
| 1  | 2  | 3  | 4  | 5 |
|----|----|----|----|---|
| 10 | 27 | 28 | 17 | 1 |
-------------------------
keine Bewertung: 5
================================================================================

================================================================================
Frage 18
--------------------------------------------------------------------------------
SmartWeb konnte meine Fragen beantworten.
keine einzige [6] praktisch alle
----------------------------
| 1  | 2  | 3  | 4 | 5 | 6 |
|----|----|----|---|---|---|
| 31 | 34 | 10 | 3 | 9 | 0 |
----------------------------
keine Bewertung: 1
================================================================================

================================================================================
Frage 19
--------------------------------------------------------------------------------
SmartWeb hat mich schnell zur gewï¿½nschten Information gefï¿½hrt.
trifft nicht zu [6] trifft zu
----------------------------
| 1  | 2  | 3 | 4 | 5 | 6 |
|----|----|---|---|---|---|
| 34 | 37 | 7 | 4 | 6 | 0 |
----------------------------
keine Bewertung: 0
================================================================================

================================================================================
Frage 20
--------------------------------------------------------------------------------
SmartWeb hat mir die richtigen Informationen geliefert.
trifft nicht zu [6] trifft zu
----------------------------
| 1  | 2  | 3  | 4  | 5 | 6 |
|----|----|----|----|---|---|
| 22 | 30 | 12 | 15 | 5 | 2 |
----------------------------
keine Bewertung: 2
================================================================================

================================================================================
Frage 21
--------------------------------------------------------------------------------
Der Gesprächsverlauf wurde eher von...
SmartWeb bestimmt [6] mir selber bestimmt
----------------------------
| 1 | 2  | 3 | 4  | 5  | 6  |
|---|----|---|----|----|----|
| 9 | 14 | 5 | 11 | 12 | 37 |
----------------------------
keine Bewertung: 0
================================================================================

================================================================================
Frage 22
```

Ich wusste immer, wie ich die nächste Eingabe (per Sprache, Tastatur, Stift)
machen konnte.
trifft nicht zu [6] trifft zu

1	2	3	4	5	6
4	5	8	10	39	22
---
keine Bewertung: 0
==========

==========
Frage 23

Ich musste meine Eingaben wiederholen.
praktisch jedes mal [6] nie

1	2	3	4	5	6
35	34	13	2	3	0
---
keine Bewertung: 1
==========

==========
Frage 24

Die Pausen zwischen Eingabe und Antwort erschienen mir ...
... sehr lang [5] ... sehr kurz

1	2	3	4	5
28	34	25	1	0
---
keine Bewertung: 0
==========

==========
Frage 25

Mit Hilfe der Spracheingabe komme ich schneller ans Ziel
trifft nicht zu [6] trifft zu

1	2	3	4	5	6
26	25	18	9	9	0
---
keine Bewertung: 1
==========

==========
Frage 26

Die Kombination von Stift- und Spracheingabe finde ich sinnvoll.
trifft nicht zu [6] trifft zu

1	2	3	4	5	6
0	0	4	9	24	51
---
keine Bewertung: 0
==========

==========
Frage 27

Im Gespräch mit SmartWeb fühlte ich mich ...
... unwohl [6] ... wohl

| 1 | 2 | 3 | 4 | 5 | 6 |

```
|---|---|----|----|----|---|
| 6 | 5 | 13 | 33 | 24 | 7 |
---------------------------
keine Bewertung: 0
================================================================================

================================================================================
Frage 28
--------------------------------------------------------------------------------
Ich bin von der Leistung von SmartWeb
... enttï¿½uscht [6] ... beeindruckt
-----------------------------
| 1  | 2  | 3  | 4  | 5 | 6 |
|----|----|----|----|---|---|
| 21 | 29 | 20 | 11 | 6 | 1 |
-----------------------------
keine Bewertung: 0
================================================================================

================================================================================
Frage 29
--------------------------------------------------------------------------------
Der Umgang mit SmartWeb hat ...
mich gelangweilt [6] mir Spaß gemacht
-----------------------------
| 1 | 2 | 3  | 4  | 5  | 6  |
|---|---|----|----|----|----|
| 9 | 6 | 24 | 20 | 15 | 14 |
-----------------------------
keine Bewertung: 0
================================================================================



### SPRACHERKENNEREVALUATION ##################################################

Die Auswertung der Spracherkennerergebnisse liefert eine Correctness von 50.63%
und eine Accuracy von 45.06% bei N=17001 (H=8608, D=4712, S=3681, I=948).

================================================================================
```

8.5 Task Level Documents

Aufgabenstellung 1 (2. Sitzung) Thema: Fußball

Stellen Sie sich vor, Sie treffen sich demnächst mit Freunden zum Fußballschauen. Um auf diesen Abend vorbereitet zu sein, informieren Sie sich mit Hilfe von SmartWeb über die Fußball-Weltmeisterschaft 2006, über die aktuellen EM-Qualifikations-Spiele und über die aktuelle 1. Bundesliga. Erkundigen Sie sich zum Beispiel nach: · Mannschaften · Titel wie DFB-Pokal, UEFA-Cup, Championsleague · schönste Tore · Torschützenkönige · erfolgreiche Fußballspieler · Siege und Niederlagen · Fairness - rote und gelbe Karten · Trainer · Vereinsgeschichte · beliebte Schiedsrichter · Spielorte

Bedingungen: Büro/Cafe, WLAN/UMTS, Headset, push-to-talk

Zusatzaufgabe: Sie haben kürzlich mit einem Freund uber Fußball diskutiert. Versuchen Sie mit Hilfe von SmartWeb zu belegen, dass Ihre Lieblingsmannschaft die sympathischere und bessere ist.

Aufgabenstellung 2 (3. Sitzung) Thema: Städteausflug

Sie wollen am Wochenende einen Ausflug in eine deutsche Stadt Ihrer Wahl unternehmen. Planen Sie Ihren Kurzurlaub mit Hilfe von SmartWeb, indem Sie sich z.B. nach folgenden Reiseinformationen erkundigen: · Anreisemöglichkeiten · Verkehrslage · Wetter · Sehenswürdigkeiten · Übernachtungsmöglichkeiten · abendliche Aktivitäten wie Kino, Theater oder Kneipen · Museen

Bedingungen: Büro/Cafe, WLAN/UMTS, Headset, ASR offen/push-to-talk

Aufgabenstellung 3 (4.Sitzung) Thema: Kinoabend

Stellen Sie sich vor, Sie möchten heute abend ins Kino gehen, wissen aber noch nicht, für welchen Film Sie sich entscheiden sollen. Benutzen Sie SmartWeb zur Meinungsfindung und holen Sie Informationen über die zur Wahl stehenden Filme ein, z.B.: · welche Filme laufen in welchen Kinos · zu welchen Uhrzeiten laufen welche Filme · Kritiken · Trailer · Inhalt · Altersbeschränkung · Reservierungsmöglichkeiten · Schauspieler · Regisseur · wie komme ich zu diesem Kino (Auto, Fußweg, öffentlich, ...) · Restaurant in der Nähe · weitere Veranstaltungen ·

Bedingungen: Büro, WLAN, Headset, ASR offen/push-to-talk

Aufgabenstellung 4 (5.Sitzung) Thema: Rundgang in Berlin

Stellen Sie sich vor, Sie wollen einen Tag in Berlin verbringen. Finden Sie heraus, was es alles zu sehen gibt und planen Sie mit Hilfe von SmartWeb eine(n) Stadtrundgang/-fahrt. Erkundigen Sie sich z.B. nach: · geführten Stadtbesichtigungen · Uhrzeiten · berühmten Bauwerken (dazu Bilder, Daten, Architekt, Anlaß, usw.) · Sehensawü rdigkeiten (fragen Sie gezielt nach Bildern · Künstlern aus Berlin · bekannten Cafes · Besuchszeiten des Bundestages · Veranstaltungen · Museen (Öffnungszeiten, Eintrittspreise, Sonderausstellungen, usw.) · Geschichte von Berlin ·

Bedingungen: Büro, WLAN, Headset, push-to-talk

Aufgabenstellung 5 (6. Sitzung) Thema: Kinoabend

Stellen Sie sich vor, Sie möchten heute abend ins Kino gehen, wissen aber noch nicht, für welchen Film Sie sich entscheiden sollen. Benutzen Sie SmartWeb zur Meinungsfindung und holen Sie Informationen über die zur Wahl stehenden Filme ein, z.B.: · welche Filme laufen in welchen Kinos · zu welchen Uhrzeiten laufen welche Filme · Kritiken · Trailer · Inhalt · Altersbeschränkung · Reservierungsmöglichkeiten · Schauspieler · Regisseur · wie komme ich zu diesem Kino (Auto, Fußweg, öffentlich, ...) · Restaurant in der Nähe · weitere Veranstaltungen ·

Bedingungen: Straße/Büro, UMTS/WLAN, Headset, push-to-talk

Aufgabenstellung 6 (7. Sitzung) Thema: vorgegebene Fragen

Sie bekommen heute eine Liste mit 15 Anfragen, die Sie der Reihe nach an SmartWeb stellen sollen. Sollte eine Anfrage im ersten Anlauf nicht zum gewünschten Ergebnis führen, sollen Sie einmal mit dem Mittel Ihrer Wahl (Sprache oder Stift) eine Korrektur vornehmen, um eine richtige Antwort zu bekommen. Bei einer passenden Antwort oder Teilantwort stellen Sie bitte eine Folgefrage.

1. Hast Du aktuelle Bilder von der Gedächtniskirche in Berlin?
2. Ich möchte Bilder von der Kongresshalle am Alexanderplatz sehen.
3. Welche Bilder gibt es von der TU Berlin?
4. Wie sieht es am Europacenter aus?
5. Wie schaut es im Studio von Sabine Christiansen aus?
6. Zeige mir von der Siegessäule Bilder.
7. Zeige mir die Maskottchen aller Wms.
8. Wer war 2002 Weltmeister?
9. Wie komme ich von Berlin nach München?
10. Ich brauche Informationen zur Verkehrslage auf der A9.
11. Wie ist das Wetter in Bamberg?
12. Was läuft momentan im Kino?
13. Wie heißt die Hauptstadt der Slowakei?
14. Wie komme ich vom Hauptbahnhof zum Brandenburger Tor?
15. Welche Hotels gibts es in der Nähe vom Flughafen?

Bedingungen: Straße(Cafe)/Büro, UMTS/WLAN, Headset, push-to-talk

Aufgabenstellung 7 (8.Sitzung) Thema: Auswahl aus 1-4

Wählen Sie in diesem Versuch ein bereits bearbeitetes Thema aus den vergangenen Sitzungen aus und stellen Sie erweiterte Anfragen zu diesem Bereich.

Thema 1: Fussballabend
Thema 2: Städteausflug
Thema 3: Kinoabend
Thema 4: Rundgang in Berlin

Bedingungen: Büro/Straße, WLAN/UMTS, Headset, ASR offen

Aufgabenstellung 8 (9.Sitzung) Thema: Arztbesuch

Sie planen einen Arztbesuch in der Umgebung. Fragen Sie SmartWeb, wo sich welche Arztpraxis befindet und wie man dort hin gelangt.Informieren Sie sich über:
verschiedene Arzte · Anreise zum Arzt · Notdienste · Apotheken

Bedingungen: Straße/Büro, UMTS/WLAN, Headset, push-to-talk

Aufgabenstellung 9 (10.Sitzung) Thema: persönliche Anfragen

Stellen SieAufgabenstellung 9 (10.Sitzung) Thema: pers250nliche Anfragen o

Stellen Sie heute Fragen, die f250r Sie von pers250nlichem Interesse sind, u o l hren Tagesablauf oder Ihr Umfeld betreffen oder stellen Sie Fragen danach, was S ie schon immer einmal wissen wollten. Greifen Sie dazu auch auf die von Ihnen vo rbereiteten Anfragen zur250ck, die Sie an u Internet-Suchmaschinen oder Sprachdi alogsysteme gestellt haben. Bedingungen: Strasse UMTS Headset push-to-talk heute Fragen, die für Sie von persönlichem Interesse sind, Ihren Tagesablauf oder Ihr Umfeld betreffen oder stellen Sie Fragen danach, was Sie schon immer einmal wissen wollten. Greifen Sie dazu auch auf die von Ihnen vorbereiteten Anfragen zurück, die Sie an Internet-Suchmaschinen oder Sprachdialogsysteme gestellt haben.

Bedingungen: Straße, UMTS, Headset, push-to-talk

8.6 Tested Conditions

The following table lists all performed evaluation session together with their most prominent conditions. The columns from left to right are: task level, session ID, connection type (WLAN/UMTS), environment (indoor/outdoor), headset type (none, MDA_Pro), ASR activation type per buttom push (ptt) or open microphone (active) and test location ('Büro' = office, Cafe, 'Straße' = street).

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location
0	e004	WLAN	indoor	MDA_Pro	ptt	Büro
0	e005	WLAN	indoor	MDA_Pro	ptt	Büro
0	e006	WLAN	indoor	MDA_Pro	ptt	Büro
0	e009	WLAN	indoor	MDA_Pro	ptt	Büro
0	e013	WLAN	indoor	MDA_Pro	ptt	Büro
0	e015	WLAN	indoor	MDA_Pro	ptt	Büro
0	e016	WLAN	indoor	MDA_Pro	ptt	Büro
0	e018	WLAN	indoor	MDA_Pro	ptt	Büro
0	e031	WLAN	indoor	MDA_Pro	ptt	Büro
0	e065	WLAN	indoor	MDA_Pro	ptt	Büro

(10 rows)

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location
1	e002	WLAN	indoor	MDA_Pro	ptt	Büro
1	e017	WLAN	indoor	MDA_Pro	ptt	Cafe
1	e019	WLAN	indoor	MDA_Pro	ptt	Cafe
1	e021	WLAN	indoor	MDA_Pro	ptt	Cafe
1	e022	WLAN	indoor	MDA_Pro	ptt	Cafe
1	e023	WLAN	indoor	MDA_Pro	ptt	Cafe
1	e025	WLAN	indoor	MDA_Pro	ptt	Büro
1	e032	WLAN	indoor	MDA_Pro	ptt	Cafe
1	e046	WLAN	indoor	MDA_Pro	active	Cafe
1	e067	WLAN	indoor	MDA_Pro	ptt	Cafe

(10 rows)

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location
2	e007	WLAN	indoor	MDA_Pro	ptt	Büro
2	e010	WLAN	indoor	MDA_Pro	ptt	Büro
2	e014	WLAN	indoor	MDA_Pro	ptt	Büro
2	e020	WLAN	indoor	MDA_Pro	ptt	Cafe
2	e024	WLAN	indoor	MDA_Pro	ptt	Büro
2	e036	WLAN	indoor	MDA_Pro	ptt	Büro
2	e047	WLAN	indoor	MDA_Pro	active	Büro
2	e048	WLAN	indoor	MDA_Pro	active	Büro
2	e071	WLAN	indoor	MDA_Pro	active	Büro

(9 rows)

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location

3	e008	WLAN	indoor	MDA_Pro	ptt	Büro
3	e011	WLAN	indoor	MDA_Pro	ptt	Büro
3	e026	WLAN	indoor	MDA_Pro	ptt	Büro
3	e029	WLAN	indoor	MDA_Pro	ptt	Büro
3	e035	WLAN	indoor	none	ptt	Büro
3	e038	WLAN	indoor	MDA_Pro	ptt	Büro
3	e039	WLAN	indoor	MDA_Pro	ptt	Büro
3	e041	WLAN	indoor	MDA_Pro	ptt	Büro
3	e044	WLAN	indoor	MDA_Pro	ptt	Büro
3	e045	WLAN	indoor	none	active	Büro

(20 rows)

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location
4	e012	WLAN	indoor	MDA_Pro	ptt	Büro
4	e027	WLAN	indoor	MDA_Pro	ptt	Büro
4	e030	WLAN	indoor	MDA_Pro	ptt	Büro
4	e033	WLAN	indoor	MDA_Pro	ptt	Büro
4	e034	WLAN	indoor	MDA_Pro	ptt	Büro
4	e040	WLAN	indoor	MDA_Pro	ptt	Büro
4	e056	WLAN	indoor	MDA_Pro	ptt	Büro
4	e057	WLAN	indoor	MDA_Pro	active	Büro
4	e075	WLAN	indoor	MDA_Pro	ptt	Büro
4	e077	WLAN	indoor	MDA_Pro	active	Büro

(10 rows)

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location
5	e050	WLAN	indoor	MDA_Pro	active	Büro
5	e051	WLAN	indoor	MDA_Pro	ptt	Büro
5	e053	WLAN	indoor	MDA_Pro	ptt	Büro
5	e054	WLAN	indoor	MDA_Pro	active	Büro
5	e055	WLAN	indoor	MDA_Pro	ptt	Büro
5	e066	WLAN	indoor	MDA_Pro	ptt	Büro
5	e069	WLAN	indoor	MDA_Pro	ptt	Büro
5	e074	WLAN	indoor	MDA_Pro	active	Büro
5	e076	WLAN	indoor	none	ptt	Büro
5	e080	UMTS	outdoor	MDA_Pro	ptt	Straße

(10 rows)

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location
6	e037	UMTS	outdoor	MDA_Pro	ptt	Straße
6	e042	UMTS	outdoor	MDA_Pro	ptt	Cafe
6	e043	UMTS	outdoor	MDA_Pro	ptt	Straße
6	e052	UMTS	outdoor	MDA_Pro	ptt	Straße
6	e061	WLAN	indoor	MDA_Pro	ptt	Büro
6	e063	UMTS	outdoor	MDA_Pro	ptt	Straße
6	e064	UMTS	outdoor	MDA_Pro	ptt	Straße
6	e072	UMTS	outdoor	MDA_Pro	ptt	Straße
6	e078	UMTS	outdoor	MDA_Pro	ptt	Straße
6	e083	UMTS	outdoor	MDA_Pro	ptt	Straße

(10 rows)

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location
7	e049	WLAN	indoor	MDA_Pro	active	Büro
7	e058	WLAN	indoor	MDA_Pro	active	Büro
7	e062	WLAN	indoor	MDA_Pro	active	Büro
7	e068	WLAN	indoor	MDA_Pro	active	Büro
7	e070	WLAN	indoor	MDA_Pro	active	Büro
7	e073	WLAN	indoor	MDA_Pro	active	Büro
7	e081	UMTS	outdoor	MDA_Pro	active	Straße
7	e084	UMTS	outdoor	MDA_Pro	active	Straße

(8 rows)

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location
8	e079	UMTS	outdoor	MDA_Pro	ptt	Straße
8	e086	WLAN	indoor	MDA_Pro	ptt	Büro
8	e087	UMTS	outdoor	MDA_Pro	ptt	Straße
8	e090	WLAN	indoor	MDA_Pro	ptt	Büro
8	e091	WLAN	indoor	MDA_Pro	ptt	Büro
8	e094	WLAN	indoor	MDA_Pro	ptt	Büro

(6 rows)

e_task_number	s_session_name	s_connection_type	s_situation	s_headsettype	s_asr_state	s_directions_location
9	e085	UMTS	outdoor	MDA_Pro	ptt	Straße
9	e089	UMTS	outdoor	MDA_Pro	ptt	Straße
9	e093	UMTS	outdoor	MDA_Pro	ptt	Straße
9	e095	UMTS	outdoor	MDA_Pro	ptt	Straße
9	e096	UMTS	outdoor	MDA_Pro	ptt	Straße

(5 rows)

8.7 Questionnaire A – Results

In the following you will find the raw results from the initial questionnaire A answered by each test subject prior to the first test. Answers with 0 hits are not shown.

```
 uea_experience_internet | count
-------------------------+-------
 ein paar Mal pro Woche  |     1
 etwa einmal pro Tag     |     1
 mehrmals am Tag         |     8
(3 rows)

              uea_experience_diasys               | count
--------------------------------------------------+-------
 Ich verwende nur selten ein Dialogsystem         |     7
 Ich verwende sie jede Woche                      |     1
 Ich verwende sie nicht mehr als einmal im Monat  |     2
(3 rows)

          uea_pre_opinion_diasys           | count
-------------------------------------------+-------
 1 ... trotzdem lieber mit einem Menschen  |     3
 2                                         |     3
 3                                         |     4
(3 rows)

        uea_pre_payfor_diasys          | count
---------------------------------------+-------
 bereit, 10 Cent pro Minute zu zahlen  |     6
 bereit, 25 Cent pro Minute zu zahlen  |     1
 nicht bereit, etwas zu zahlen         |     3
(3 rows)

 uea_pre_opinion_service | count
-------------------------+-------
 Stimme ich voll zu      |     1
 Tendenziell richtig     |     4
 Weitgehend richtig      |     5
(3 rows)

 uea_pre_opinion_human | count
-----------------------+-------
 Stimme ich voll zu    |     3
 Tendenziell falsch    |     1
 Tendenziell richtig   |     3
 Weitgehend falsch     |     1
 Weitgehend richtig    |     2
(5 rows)

 uea_pre_opinion_personalassi | count
------------------------------+-------
 Stimme ich voll zu           |     2
 Tendenziell falsch           |     1
 Tendenziell richtig          |     3
 Weitgehend falsch            |     1
 Weitgehend richtig           |     3
(5 rows)

   uea_pre_opinion_help   | count
--------------------------+-------
 Finde ich extrem wichtig |     6
 Ziemlich wichtig         |     4
(2 rows)

 uea_synthesis_simple | count
----------------------+-------
```

```
 einige Male          |     1
 oft                  |     2
 sehr oft             |     5
 selten               |     2
(4 rows)

 uea_synthesis_sms | count
-------------------+-------
 einige Male       |     5
 nie               |     2
 oft               |     1
 selten            |     2
(4 rows)

 uea_synthesis_nav_simple | count
--------------------------+-------
 einige Male              |     5
 nie                      |     2
 oft                      |     2
 selten                   |     1
(4 rows)

 uea_synthesis_nav_complex | count
---------------------------+-------
 einige Male               |     7
 nie                       |     1
 oft                       |     1
 selten                    |     1
(4 rows)
```

8.8 Questionnaire C – Results

In the following you will find the raw results from the initial questionnaire C answered by each test subject right after the last test. Answers with 0 hits are not shown.

```
        uec_post_payfor_diasys        | count
--------------------------------------+-------
 bereit, 10 Cent pro Minute zu zahlen |     3
 bereit, 25 Cent pro Minute zu zahlen |     2
 nicht bereit, etwas zu zahlen        |     5
(3 rows)

 uec_post_opinion_service | count
--------------------------+-------
 Stimme ich voll zu       |     2
 Tendenziell richtig      |     3
 Weitgehend falsch        |     1
 Weitgehend richtig       |     4
(4 rows)

 uec_post_opinion_human | count
------------------------+-------
 Stimme ich voll zu     |     3
 Tendenziell falsch     |     1
 Tendenziell richtig    |     3
 Weitgehend richtig     |     3
(4 rows)

 uec_post_opinion_personalassi | count
```

```
------------------------------+-------
 Stimme ich voll zu           |     1
 Stimme ich überhaupt nicht zu |     3
 Tendenziell falsch           |     2
 Tendenziell richtig          |     3
 Weitgehend richtig           |     1
(5 rows)

 uec_post_opinion_help | count
-----------------------+-------
 Stimme ich voll zu    |     6
 Tendenziell falsch    |     1
 Weitgehend richtig    |     3
(3 rows)

 uec_voice_fit_to_system | count
-------------------------+-------
 Gut                     |     3
 Ordentlich              |     5
 Schlecht                |     2
(3 rows)

 uec_voice_quality | count
-------------------+-------
 Gut               |     1
 Ordentlich        |     5
 Schlecht          |     4
(3 rows)

 uec_voice_pleasentness | count
------------------------+-------
 Angenehm               |     2
 Neutral                |     3
 Sehr angenehm          |     1
 Sehr unangenehm        |     1
 Unangenehm             |     3
(5 rows)

 uec_voice_naturalness | count
-----------------------+-------
 Natürlich             |     1
 Neutral               |     3
 Sehr unnatürlich      |     2
 Unnatürlich           |     4
(4 rows)

 uec_voice_applicability | count
-------------------------+-------
 Ja                      |     5
 Nein                    |     5
(2 rows)

 uec_additional_questions | count
--------------------------+-------
 Stimme ich voll zu       |     9
 Weitgehend richtig       |     1
(2 rows)

 uec_defined_formulation | count
-------------------------+-------
 Stimme ich voll zu      |     1
 Tendenziell falsch      |     2
```

```
 Tendenziell richtig     |     2
 Weitgehend falsch       |     1
 Weitgehend richtig      |     4
(5 rows)

  uec_know_infotype  | count
---------------------+-------
 Stimme ich voll zu  |     3
 Tendenziell falsch  |     2
 Tendenziell richtig |     1
 Weitgehend richtig  |     4
(4 rows)

  uec_acknowledge   | count
--------------------+-------
 Stimme ich voll zu |     6
 Weitgehend richtig |     4
(2 rows)

   uec_processing   | count
--------------------+-------
 Stimme ich voll zu |    10
(1 row)

  uec_modify_input  | count
--------------------+-------
 Stimme ich voll zu |    10
(1 row)
```