Human-Computer Interaction

Session 7:
User Interface Evaluation

Reading:
- Dix et al., Human-Computer Interaction, chapter 9
- Shneiderman, Designing the User Interface, chapter 4

Usability (ISO 9241)

Usability = The effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments.

Effectivity
- Accuracy and completeness with which the users can in principle achieve a specific goal.

Efficiency
- Effort expended in relation to the accuracy and completeness (quality) of the achieved results

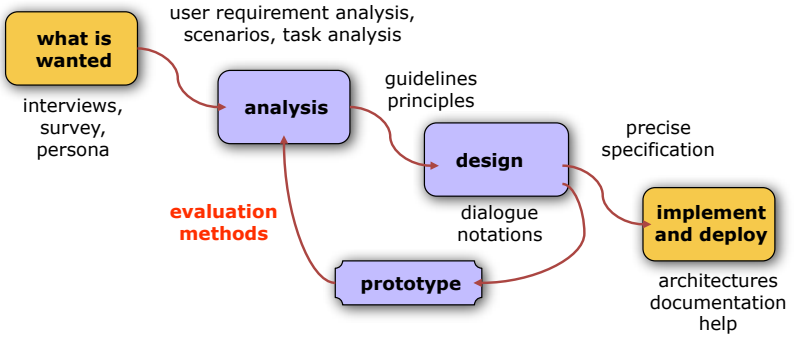
Satisfaction
- Positive attitude of the user towards using the system
- Freedom of using the system without restrictions

Methods in user-centered design

1. Field studies
2. User requirement analysis
3. Iterative design
4. Usability evaluation
5. Task analysis
6. Focus groups
7. Heuristic evaluation
8. User interviews
9. Surveys
10. …

Ranking based on a survey among experienced UCD practitioners (103 questionnaires) (Mao et al., 2005)

User-centered design process



Process to develop interactive systems such that usability will be maximized.

Prototyping

The earlier a prototype, the better

Horizontal vs. vertical prototypes

- horizontal: complete interface, no/little function
- vertical: functions (partially) implemented
- mixtures of both useful and common

Stages of prototyping

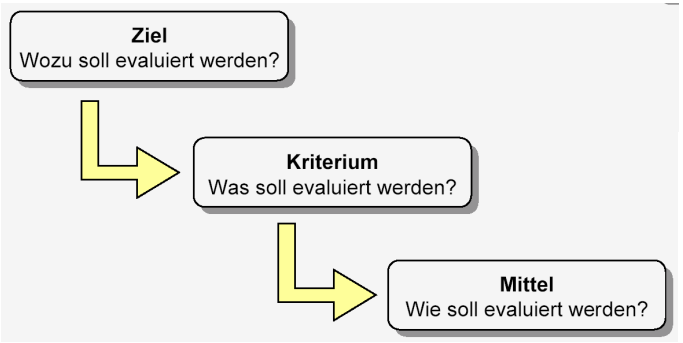
- conceptual prototype: description/spec and imagines of how the system is about to work
- paper prototype: sketches, drafts, pictures, etc.
- static screens: single screen design snapshots
- dynamic simulation: simulations of simple procedures
- Wizard-of-Oz: operated by invisible person („wizzard")

Key questions for today

How can the usability of a system be evaluated?

How can usability problems be found and improvements suggested?

Evaluation = Testing to what degree a system adheres to previously defined criteria



Key questions for an evaluation

Why? assess usability and user effects, find problems, make suggestions for improvement

What? lay down usability criteria

Where? in the lab or in the field

Who? experts (with/without user) or real users

When? in all design stages (concept, prototypes, impl.)

- Summative evaluation: final quantitative assessment of initially defined criteria
- Formative evaluation: at different times, assess current system against actual requirements

Evaluation procedure

1. Define criteria for the system to be usable
2. Define observables and performance levels for each criterion („operationalization")
3. Measurement (Analysis)
   - application of criteria and comparison with performance levels
4. Assessment (Synthesis)
   - make judgement based on results
   - derive suggestions for improvement on the criteria

Choosing methods and design

Validity (*Gültigkeit*): will criteria be observed/measured?

Reliability (*Zuverlässigkeit*): is the study reproducible?

Significance and Generalisation (aka. external validity): Selection of participants, influence of the context of the study on observed behavior?

Pilot/Pre-Study
- if something is not fully clear, always make a pre-study
- test feasibility and practicability, practice procedure, improve
- can employ colleagues as test subjects
- a row of pre-studies might possibly be required

Evaluation methods

Usability inspection (expert reviews)
- Guidelines review & consistency inspection
- Cognitive walkthrough
- Heuristic evaluation
- Focus group

User studies
- Usability testing
- Thinking-Aloud
- Field studies
- Interviews & questionnaires

Model-based evaluation

Usability inspection methods

Guidelines Review
Consistency Inspection
Cognitive Walkthrough
Heuristic Evaluation

Guideline review & consistency inspection

Guideline review

- expert checks interface for conformance with guidelines, either standard guidelines, e.g. Shneiderman's rules, or organization-specific guidelines, e.g. styleguide

Consistency inspection

- expert checks interface for consistency of terminology, colors, fonts, icons, menues, general layouts, etc.
- within interface as well as documentation, training material, online help

Cognitive Walkthrough

Task-oriented inspection method
(„Benutzbarkeits-Gedankenexperiment")

Expert simulates user walking through the interface to carry out typical tasks

- select task and perform it step by step
- select all relevant tasks, simulate day in the life of the user
- can identify potential problems for a user

Advantage:

- Can be carried out and spot mis-conceptions early on

Problem:

- Can an evaluator ever „simulate" a user? May also employ users as evaluators

Cognitive Walkthrough

1. Prepration

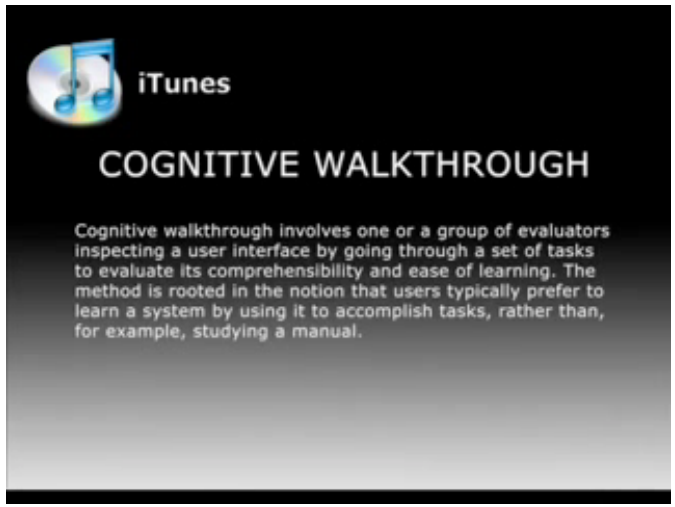
- Detailed spec of potential user
- Detailed spec of task, structured in single steps
- List of possible actions and their results
- Prototype of the system (paper, partially implemented, etc.)

2. Analysis

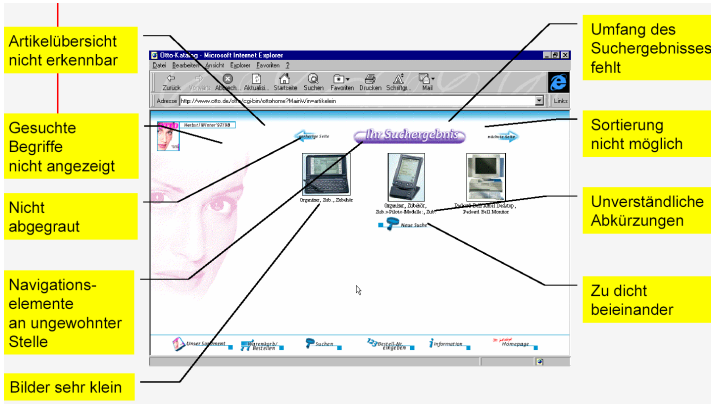
- Expert walks through all actions and system responses, each time answering the following questions:
  - Are the right actions available (effects = user goals/intentions)?
  - Will the user be able to identify the actions as such?
  - Will the user find the correct actions?
  - Will the user understand the system feedback?

3. **Follow-Up**

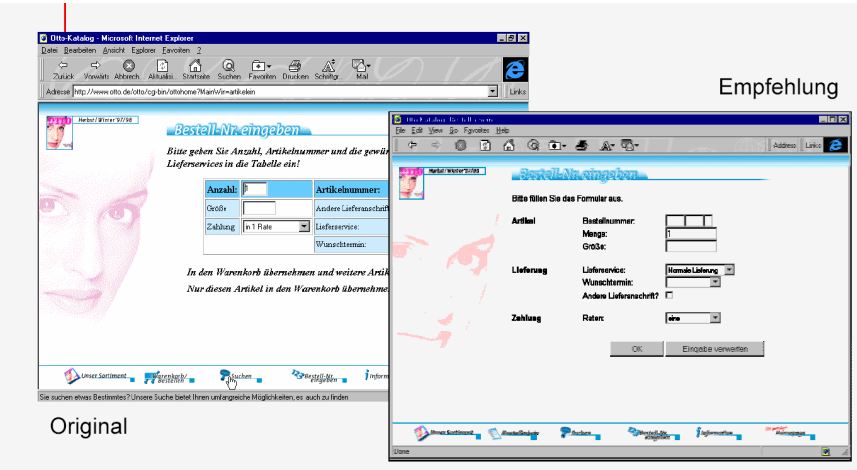
- Recordings of results and ideas about alternative design and further improvements



Example: inspection of Otto Versand webpage...



...and recommendations



Heuristic Evaluation

J. Nielsen (1993)
www.useit.com

Experts critique an interface (either system or running prototype) to determine conformance with a short list of general design heuristics

Can and should be conducted by multiple experts independently (interface developer or usability experts)

Check heuristics/design rules, e.g.:

- Shneiderman's 8 golden rules of interface design
- Nielsen's 10 heuristics (1993; cf. previous session)
- Extended heuristics as of 2001 (Nielsen, 2001)

Usability heuristics (1)

Visibility of system status

Match between system and the real world

- Speak the users' language, follow real-world conventions, make information appear in a natural and logical order

User control and freedom

- Provide a clearly marked "emergency exit" to leave an unwanted state (undo and redo)

Consistency and standards

- Users should not have to wonder whether different words, situations, or actions mean the same thing.

Error prevention

Usability heuristics (2)

Recognition rather than recall

Flexibility and efficiency of use

- cater both inexperienced and experienced users, allow to tailor frequent actions

Aesthetic and minimalist design

- provide no irrelevant or rarely needed info

Help users recognize, diagnose, and recover from errors

- Error messages in plain language (no codes), precisely indicate the problem, suggest a solution.

Help and documentation

- provide help and documentation, easy to search, focus on user task, list concrete steps to be carried out, not too large

Heuristic Evaluation

1. **Training session**
   - Reviewers practice detailed heuristics

2. **Evaluation**
   - Each reviewer evaluates with a list of standard heuristics the interface - normally 4 iterations
   - Tests the general flows of tasks and functions of the various interface elements (not strictly task-oriented)
   - Observer takes notes of identified problems
   - Reviewers communicate only after their iterations

Heuristic Evaluation

3. **Results and reviewer session**
   - Make list of problems (violated principles+reasons)
   - Detailed descriptions of the problems

4. **Problem assessment**
   - How serious and unavoidable is a usability problem?
   - Each reviewer assesses each identified problem with respect to its severity:
     - 0 - don‘t agree that this is a usability problem
     - 1 - cosmetic problem
     - 2 - minor usability problem
     - 3 - major usability problem - important to fix
     - 4 - usability catastrophe; imperative to fix
   - Final ranking of all problems

Heuristic Evaluation

Example:

- *Interface used command „Save“ on 1st screen for saving the user‘s file, but used „write file“ on 2nd screen. Users may be confused by this different terminology.*
- Violation of consistency/standards - severity rating 3

Advantage:

- fast, cheap, qualitatively good results

Problems:

- experts aren‘t real users
- heuristics do not cover all possible problems

Example:
outcome
evaluation form

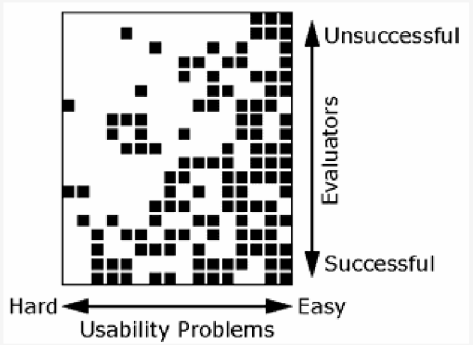
Heuristic Evaluation Form

Name: Daniel Robinson

Severity Scale:
0 - don't agree that this is a usability problem
1 - cosmetic problem
2 - minor usability problem
3 - major usability problem; important to fix
4 - usability catastrophe; imperative to fix

	How found?	Heuristic violated	Problem description and how heuristic is violated	Severity	Solutions
1	Browsing the entitlements page/table.	Consistency and standards	There are no units for the price/acre on the entitlements table/page and currency isn't specified either. For example, it says 300 per acre for stone wall protection and management rather than £300 per acre, or €300 per acre etc.	2	Display the currency (£, €, $US etc.) by the values, in the column header, or somewhere else on the page.
2	Noticed a horizontal scroll bar on the web browser window.	Recognition rather than recall	Admin Home was so far over to the right in the menu bar at the bottom, that I had to resize the window to see it.	2	Wrap the text underneath, or make site variable width or something.
3	After using the site for a while, it was eventually noticed.	Consistency and Standards	Menu bar is at the bottom of the page (in fact, I didn't notice it for about the first 5 minutes of using the site). Menus normally appear at the top of a page or screen which would feel more natural.	2	Put menu bar at top of page instead of bottom.
4	When	Recognitio	When viewing one of the	2	Put menu bar

Wieviele Reviewer ?



19 Reviewer sollten
16 Fehler in einer
Bankensoftware finden:

- Unterschiedliche Reviewer fanden durchaus unterschiedliche Fehler
- Die erfolgreichsten Reviewer finden nicht immer die schwierigsten Fehler

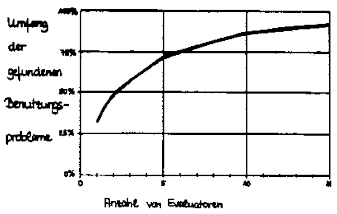
Einsatz mehrerer Reviewer sinnvoll !

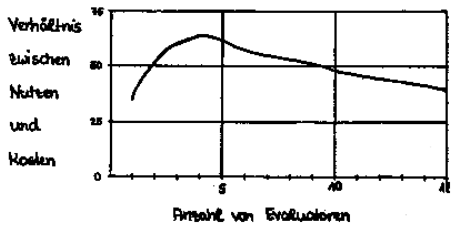
Nielsen, J.: How to conduct a heuristic evaluation, http://www.useit.com/papers/heuristic/heuristic_evaluation.html

How many expert reviewers?

Good choice: 4-5 reviewers

- Use 62 times higher than costs
- spot ~75-80% of the problems

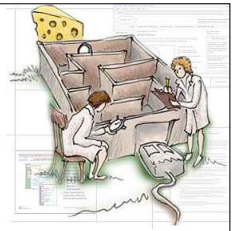




User studies

Thinking aloud
Cooperative evaluation
Interviews & questionnaires
Usability testing

User studies



In general:

Evaluate interactions between actual users and a system

Measure performance on typical tasks, for which the system was designed

Use video and interaction logging to capture errors and frequencies and time of commands, or protocols

Can be performed in the lab or the field

Users may be interviewed or complete questionnaires, to gather data about opinions, attitudes, etc.

Lab studies

- Experiment under controlled conditions
  - specialist equipment available
  - uninterrupted environment
- Disadvantages:
  - lack of context
  - difficult to observe user cooperation
- Prevalent paradigm in exp. psychology

Field studies

- Experiments dominated by group formation
- Field studies more realistic
  - *distributed cognition* ⇒ work studied in context
  - real action is *situated*
  - physical *and* social environment crucial
- sociology and anthropology – open study and rich data

Thinking Aloud

User is observed while performing a predefined task and asked to describe what ...

- s/he is expecting to happen
- s/he is thinking is happening

- Advantages
  - simplicity - requires little expertise
  - can provide useful insight into user's mental model
  - can show how system is actually used
- Disadvantages
  - artificial test situation → cooperative evaluation
  - subjective and selective → multiple trials & users needed
  - act of describing may alter task performance

Cooperative Evaluation

- User evalutes together with expert,
  - sees himself as collaborator
  - both can ask each other questions
- Additional advantages
  - less constrained and easier to use
  - user is encouraged to criticize system
  - clarification dialogues possible
- Problems with *both* techniques
  - generate a large volume of information (*protocols*)
  - 'Protocol analysis' crucial and time-consuming

Query techniques

Interviews:

- analyst questions user, based on prepared questions
- pro: relatively cheap, issues can be explored more fully, can reveal unanticipated problems
- contra: informal, subjective, can be suggestive

Questionnaires:

- fixed questions given to users
- style of questions: open vs. closed, scalar vs. binary, multiple-choice, ordering, negative vs. positive, ...
- style of answers: text, yes/no, number of options, ...
- pro: reaches large user group, can be analyzed rigorously, applicable when interactions themselves can or should not be monitored
- contra: need careful design, less flexible, less probing

Several standard questionnaires available

Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology

By: Fred D. Davis
Computer and Information Systems
Graduate School of Business Administration
University of Michigan
Ann Arbor, Michigan 48109

AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität[1]

Marc Hassenzahl
Technische Universität Darmstadt

Michael Burmester
Hochschule der Medien Stuttgart

Franz Koller
User Interface Design GmbH Ludwigsburg

Usability Testing

- observe and record user behavior under typical situations and tasks
  - video, audio
  - mouse & keyboard logging
  - eye gaze
- use data to calculate processing time, find common user errors, understand why users behave like that
- evaluate subjective “satisfaction” by means of additional questionnaires or interviews

Usability Testing vs.	**Controlled Experiment**
few users	many users to have sufficient data for statistics
designed to find flaws in interface design	designed to show statistically significant differences between conditions (hypotheses)
outcome: report with recommended changes	outcome: validation or rejection of a hypothesis
carefully designed task	carefully designed task

Usability Testing

1. get representative users
   - 5-10 participants
2. define criteria for evaluation, e.g.:
   - time for task completion
   - time for task after distraction/new input
   - number and kind of errors per task and unit time
   - number of access to online help or manual
   - ...
3. develop test scenario: setup + context + task
   - choose relevant scenarios (typical vs. extreme)
   - keep task duration shorter than 30 minutes
   - ensure identical conditions for all participants
4. consider ethical issues
   - de-brief participants, get consent, etc.

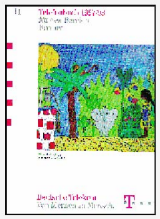
Usability Testing

4. run pilot tests & refine design
   - pratice with staff and observers
5. actual testing
   - instruction of participants
   - carry out test and record data
6. analysis
   - statistics, e.g. mouse events, menue selection
   - screen design: gaze tracking and course of task completion
   - post task video confrontation and user interview
7. report results and make recommendations for improvement

Usability Testing - Example

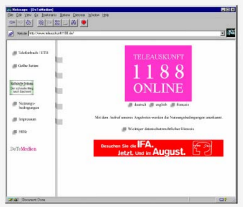
Ziel: Vergleich unterschiedlicher Telefonauskunftsysteme

- hinsichtlich ihrer Benutzbarkeit
- Verfahren: Vier Versuchspersonen bearbeiten jeweils 4 Prüfaufgaben.
- Die Bearbeitung wird mit Video, Audio und Logging-Programmen protokolliert.

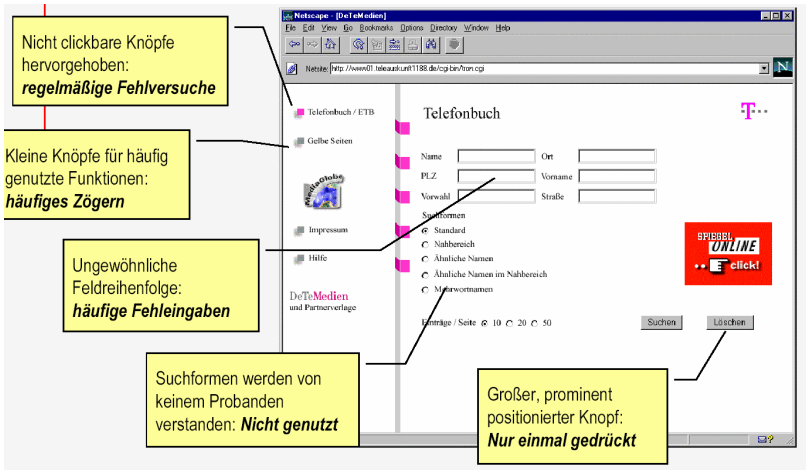


Telefonbuch

Telefon-CD
der DeTeMedien

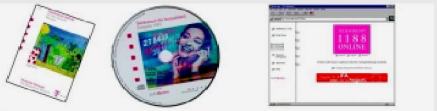


www.teleauskunft1188.de



Zeitdauer & Korrektheit im Vergleich

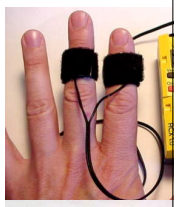
Zusammengefaßte Ergebnisse

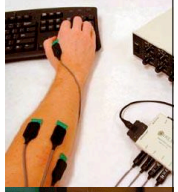


Aufgabenstellung				
1. Suche die Telefonnummer von Maria Müller. Sie wohnt Am Ziegelberg in Bremen.	Korrekte Ergebnisse	★★★	★★	★★★★★
	Bearbeitungs-dauer [min]	0:45	2:30	3:00
2. Suche die private Telefon-nummer von Carsten Bormann (TZI-Bereich Digitale Medien und Netze).	Korrekte Ergebnisse	★	★	
	Bearbeitungs-dauer [min]	0:30	1:00	2:45
3. Marc-Oliver Schulze wohnt bei seinem Vater in Bremen. Seine Telefonnummer beginnt mit einer "40".	Korrekte Ergebnisse	★★★★★	★★★★	
	Bearbeitungs-dauer [min]	1:15	1:50	4:10
4. Suche einen Sportarzt in Bremen.	Korrekte Ergebnisse	★★★	★	★½
	Bearbeitungs-dauer [min]	0:30	2:30	4:20

Beobachtung
Usability Test

Physiological measurements





May help determine a user's reaction to an interface (emotion, arousal, stress, fatigue, ...)

measurements include:

- heart activity, including blood pressure and pulse
- activity of sweat glands: Galvanic Skin Response
- electrical activity in muscle: electromyogram
- electrical activity in brain: electroencephalogram
- ...

Difficult to interpret physiological responses

Eye tracking

Eye movement and gaze patterns reflect amount of cognitive processing a display requires

Measurements include

- fixations: eye maintains stable position. number and duration indicate level of difficulty with display (`heat maps´)
- saccades: rapid eye movement from one point of interest to another
- scan paths: moving straight to a target with a short fixation at the target is optimal