Reminder

Requirements for Successful Completion

- completion of practical project
- participation in user studies
- written report (graded)

Appointment for Participation in User Study

- please have a look at the document „Usability_Study_Food_Assist.pdf“
- book for your appointment via: **https://doodle.com/bp/hiteshdhiman/foodassist-studie**

Usability Engineering

Evaluation Techniques

Carsten Röcker
inIT - Institut für industrielle Informationstechnik
carsten.roecker@th-owl.de

Overview

- Introduction
- Evaluation through Expert Analysis
- Evaluation through User Participation
- Choosing an Evaluation Method
- Summary

Introduction

Goals of Evaluation

- assess extent of system functionality
- assess effect of interface on user
- identify specific problems

Introduction

Evaluation

- tests usability and functionality of system
- occurs in laboratory, field and/or in collaboration with users
- evaluates both design and implementation
- should be considered at all stages in the design life cycle

Overview

- Introduction
- **Evaluation through Expert Analysis**
  - **Cognitive Walkthrough**
  - Heuristic Evaluation
  - Review-Based Evaluation
- Evaluation through User Participation
- Choosing an Evaluation Method
- Summary

Cognitive Walkthrough

- evaluates design on how well it supports user in learning task
- usually performed by expert in cognitive psychology
- expert ‘walks though’ design to identify potential problems using psychological principles
- forms used to guide analysis

Cognitive Walkthrough

- for each task walkthrough considers
  - what impact will interaction have on user?
  - what cognitive processes are required?
  - what learning problems may occur?
- analysis focuses on goals and knowledge
  - does the design lead the user to generate the correct goals?

Cognitive Walkthrough

Guided by 4 Questions (Wharton et al., 1994)

- Will the user try to achieve the effect that the subtask has?
  - Does the user understand that this subtask is required?
- Will the user notice that the correct action is available?
  - e.g., is the button visible?
- Will the user understand that the wanted subtask can be achieved by the action?
  - e.g., the right button is visible, but the user does not understand the text and will therefore not click on it
- Will the user get feedback?
  - Does the user know that they have done the right thing after performing the action?

Overview

- Introduction
- Evaluation through Expert Analysis
  - Cognitive Walkthrough
  - **Heuristic Evaluation**
  - Review-Based Evaluation
- Evaluation through User Participation
- Choosing an Evaluation Method
- Summary

Heuristic Evaluation

- usability criteria (heuristics) are identified
- design examined by experts to see if these are violated
- example heuristics
  - system behavior is predictable
  - system behavior is consistent
  - feedback is provided
- heuristic evaluation ‘debugs‘ design

Heuristic Evaluation

Example: 10 Heuristics for UI Design (Nielsen, 1995)

1. Visibility of System Status

- users should always be informed about what is going on
- appropriate feedback within reasonable time

2. Match between System and Real World

- the system should speak the users' language
  - words, phrases and concepts should be familiar to the user
- avoid system-oriented terms
- follow real-world conventions
  - information appear in a natural and logical order

Heuristic Evaluation

3. User Control and Freedom

- users often choose system functions by mistake
- clearly marked “emergency exit“ to leave the unwanted state
- support undo and redo

4. Consistency and Standards

- users should not have to wonder whether different words, situations, or actions mean the same thing
- follow platform conventions

Heuristic Evaluation

5. Error Prevention

- even better than good error messages is a careful design, which prevents a problem from occurring in the first place
- either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action

6. Recognition rather than Recall

- minimize the user's memory load by making objects, actions, and options visible
- the user should not have to remember information from one part of the dialogue to another
- instructions for use of the system should be visible or easily retrievable whenever appropriate

Heuristic Evaluation

7. Flexibility and Efficiency of Use

- accelerators (unseen by the novice user) may often speed up the interaction for the expert user
  - system can cater to both inexperienced and experienced users
- allow users to tailor frequent actions

8. Aesthetic and Minimalist Design

- dialogues should not contain information which is irrelevant or rarely needed
  - every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility

Heuristic Evaluation

9. Help Users Recognize, Diagnose, and Recover from Errors

- error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution

10. Help and Documentation

- ideal case: system can be used without documentation
- reality: often necessary to provide help and documentation
- any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large

Heuristic Evaluation

Evaluation Process (Nielsen, 1995)

- evaluator assesses system and notes violations of heuristics, which could indicate a potential usability problem
- evaluator also assesses severity of each problem based on four factors
  - how common is the problem?
  - how easy is it for the user to overcome?
  - will it be a one-off problem or a persistent one?
  - how seriously will the problem be perceived?

Heuristic Evaluation

Overall Severity Rating (Nielsen)

- individual assessments are combined to an overall severity score for each usability problem:

0 = I don't agree that this is a usability problem at all

1 = cosmetic problem only: need not be fixed unless extra time is available on project

2 = minor usability problem: fixing this should be given low priority

3 = major usability problem: important to fix, so should be given high priority

4 = usability catastrophe: imperative to fix this before product can be released

Overview

- Introduction
- Evaluation through Expert Analysis
  - Cognitive Walkthrough
  - Heuristic Evaluation
  - **Review-Based Evaluation**
- Evaluation through User Participation
- Choosing an Evaluation Method
- Summary

Review-Based Evaluation

- results from the literature used to support or refute parts of design
- care needed to ensure results are transferable to new design
- design rationale can also provide useful evaluation information

Overview

- Introduction
- Evaluation through Expert Analysis
- **Evaluation through User Participation**
  - **Styles of Evaluation**
  - Experimental Evaluation
  - Observational Techniques
  - Query Techniques
  - Physiological Methods
- Choosing an Evaluation Method
- Summary

Styles of Evaluation

Laboratory Studies

- Advantages
  - specialist equipment available
  - uninterrupted environment
- Disadvantages
  - lack of context
  - difficult to observe several users cooperating
- Appropriate
  - if system location is dangerous
  - if system is impractical for constrained single use
  - to allow controlled manipulation of use

Styles of Evaluation

Field Studies

- Advantages
  - natural environment
  - context retained (though observation may alter it)
- Disadvantages
  - distractions
  - noise
- Appropriate
  - where context is crucial for validity

Overview

- Introduction
- Evaluation through Expert Analysis
- **Evaluation through User Participation**
  - Styles of Evaluation
  - **Experimental Evaluation**
  - Observational Techniques
  - Query Techniques
  - Physiological Methods
- Choosing an Evaluation Method
- Summary

Experimental Evaluation

- controlled evaluation of specific aspects of interactive behavior
- evaluator chooses hypothesis to be tested
- a number of experimental conditions are considered, which differ only in the value of some controlled variable
- changes in behavioral measure are attributed to different conditions

Experimental Evaluation

Experimental Factors

- Subjects
  - who (representative, sufficient sample)
- Variables
  - things to modify and measure
- Hypothesis
  - what you'd like to show
- Experimental Design
  - how you are going to do it

Experimental Evaluation

Variables

- Independent Variable (IV)
  - characteristic changed to produce different conditions
  - e.g., interface style, number of menu items
- Dependent Variable (DV)
  - characteristics measured in the experiment
  - e.g., time taken, number of errors

Experimental Evaluation

Hypothesis

- prediction of outcome
  - framed in terms of IV and DV
  - e.g., “error rate will increase as font size decreases”

- null hypothesis
  - states no difference between conditions
  - aim is to disprove this
  - e.g., null hypothesis = “no change with font size”

Experimental Evaluation

Experimental Design

- Within Groups Design
  - each subject performs experiment under each condition
  - transfer of learning possible
  - less costly and less likely to suffer from user variation
- Between Groups Design
  - each subject performs under only one condition
  - no transfer of learning
  - more users required
  - variation can bias results

Experimental Evaluation

Analysis of Data

- before you start to do any statistics:
  - look at data
  - save original data
- choice of statistical technique depends on:
  - type of data
  - information required
- type of data
  - discrete (finite number of values)
  - continuous (any value)

Experimental Evaluation

Analysis: Types of Tests

- parametric
  - assume normal distribution
  - robust
  - powerful
- non-parametric
  - do not assume normal distribution
  - less powerful
  - more reliable
- contingency table
  - classify data by discrete attributes
  - count number of data items in each group

Experimental Evaluation

Analysis of Data

- what information is required?
  - is there a difference?
  - how big is the difference?
  - how accurate is the estimate?
- parametric and non-parametric tests mainly address first of these

Overview

- Introduction
- Evaluation through Expert Analysis
- **Evaluation through User Participation**
  - Styles of Evaluation
  - Experimental Evaluation
  - **Observational Techniques**
  - Query Techniques
  - Physiological Methods
- Choosing an Evaluation Method
- Summary

Observational Techniques

- Think Aloud
- Cooperative Evaluation
- Protocol Analysis
- Automated Analysis
- Post-Task Walkthroughs

Observational Techniques

Think Aloud

- user observed performing task
- user asked to describe what he is doing and why, what he thinks is happening etc.
- advantages
  - simplicity (requires little expertise)
  - can provide useful insight
  - can show how system is actually used
- disadvantages
  - subjective
  - selective
  - act of describing may alter task performance

Observational Techniques

Cooperative Evaluation

- variation on think aloud
- user collaborates in evaluation
- both user and evaluator can ask each other questions throughout
- additional advantages
  - less constrained and easier to use
  - user is encouraged to criticize system
  - clarification possible

Observational Techniques

Protocol Analysis

- paper and pencil
  - cheap, limited to writing speed
- audio
  - good for think aloud, difficult to match with other protocols
- video
  - accurate and realistic, needs special equipment, obtrusive
- computer logging
  - automatic, unobtrusive, large amounts of data, difficult to analyze
- user notebooks
  - coarse and subjective, useful insights, good for longitudinal studies

Observational Techniques

Protocol Analysis

- mixed use in practice
- audio/video transcription difficult and requires skill
- some automatic support tools available

Observational Techniques

Post-Task Walkthroughs

- transcript played back to participant for comment
  - immediately => fresh in mind
  - delayed => evaluator has time to identify questions
- useful to identify reasons for actions and alternatives considered
- necessary in cases where think aloud is not possible

Overview

- Introduction
- Evaluation through Expert Analysis
- **Evaluation through User Participation**
  - Styles of Evaluation
  - Experimental Evaluation
  - Observational Techniques
  - **Query Techniques**
  - Physiological Methods
- Choosing an Evaluation Method
- Summary

Query Techniques

Interviews

- analyst questions user on one-to-one basis
  - usually based on prepared questions
- informal, subjective and relatively cheap
- advantages
  - can be varied to suit context
  - issues can be explored more fully
  - can elicit user views and identify unanticipated problems
- disadvantages
  - very subjective
  - time consuming

Query Techniques

Questionnaires

- set of fixed questions given to users
- advantages
  - quick and reaches large user group
  - can be analyzed more rigorously
- disadvantages
  - less flexible
  - less probing

Query Techniques

Questionnaires

- need careful design
  - what information is required?
  - how are answers to be analyzed?
- styles of question
  - general
  - open-ended
  - scalar
  - multi-choice
  - ranked

Overview

- Introduction
- Evaluation through Expert Analysis
- **Evaluation through User Participation**
  - Styles of Evaluation
  - Experimental Evaluation
  - Observational Techniques
  - Query Techniques
  - **Physiological Methods**
- Choosing an Evaluation Method
- Summary

Physiological Methods

Eye Tracking

- head or desk mounted equipment tracks the position of the eye
- eye movement reflects the amount of cognitive processing a display requires
- measurements include
  - *fixations:* eye maintains stable position, number and duration indicate level of difficulty with display
  - *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

Physiological Methods

Physiological Measurements

- emotional response linked to physical changes
- these may help determine a user's reaction to an interface
- measurements include:
  - heart activity, including blood pressure, volume and pulse
  - activity of sweat glands: Galvanic Skin Response (GSR)
  - electrical activity in muscle: electromyogram (EMG)
  - electrical activity in brain: electroencephalogram (EEG)
- some difficulty in interpreting these physiological responses (more research needed)

Overview

- Introduction
- Evaluation through Expert Analysis
- Evaluation through User Participation
  - Styles of Evaluation
  - Experimental Evaluation
  - Observational Techniques
  - Query Techniques
  - Physiological Methods
- **Choosing an Evaluation Method**
- Summary

Choosing an Evaluation Method

Decision depends on different factors:

- *when in process* (design vs. implementation)
- *style of evaluation* (laboratory vs. field)
- *how objective* (subjective vs. objective)
- *type of measures* (qualitative vs. quantitative)
- *level of information* (high level vs. low level)
- *level of interference* (obtrusive vs. unobtrusive)
- *resources available* (time, subjects, equipment, etc.)

Summary

- Introduction
- Evaluation through Expert Analysis
- Evaluation through User Participation
- Choosing an Evaluation Method
- Summary

Summary

- **Evaluation** tests the usability, functionality and acceptability of an interactive system.
- Evaluation may take place:
  - in the **laboratory** or
  - in the **field**.
- Some approaches are based on **expert evaluation**:
  - analytic methods
  - observational methods
  - query methods.

Summary

- Some approaches **involve users**:
  - experimental methods
  - observational methods
  - query methods.
- An evaluation method must be chosen carefully and must be suitable for the job.