Usability testing
Overview

• What is usability testing
• Usability testing versus traditional research
• Expert based testing
• User-based testing
• Practicality of usability testing
Warm-up exercise

• Two students together
• One student has forgotten a book at a bus number 100 and should report this to Ruter
• Another student should measure the time, evaluate the solution and find out if this was difficult
What is usability?

"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." (ISO 9241-11)
Usability dimensions

– **Effectiveness** - can users complete tasks, achieve goals with the product, i.e. do what they want to do?

– **Efficiency** - how much effort do users require to do this? (Often measured in time)

– **Satisfaction** – what do users think about the products ease of use?
....which are affected by

– The **users** - who is using the product? e.g. are they highly trained and experienced users, or novices?
– Their **goals** - what are the users trying to do with the product - does it support what they want to do with it?
– The usage situation (or 'context of use') - where and how is the product being used?
Usability cont.

• Usability is composed of:
  – **Learnability**: How easy is it for users to accomplish basic tasks the first time they encounter the design?
  – **Efficiency**: Once users have learned the design, how quickly can they perform tasks?
  – **Memorability**: When users return to the design after a period of not using it, how easily can they re establish proficiency?
  – **Errors**: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
  – **Satisfaction**: How pleasant is it to use the design? (Nielsen, 1993)
What is usability testing?

• Involving representative users, representative tasks, representative environments
  – Testing paper prototypes
  – Wizard of Oz
  – Testing working version of a system before it is released
  – Testing working system
  – Different devices (smart phones, laptops...)
Goal of usability testing

• Improve the quality of an interface by finding flaws in it
  – Flaws that cause problems for the majority of people (not preferences)
• What works fine (keep it)
What is wrong here? (groups of 5 students, 2-3 minutes)
Scope of usability testing

• Usability testing is a general term that could mean:
  – Having hundreds of users test interfaces, in different treatment groups
  – Sitting next to a total of 3 users and watching as they attempt different tasks
  – The reality is usually somewhere in between

• The best usability testing is the one that actually takes place
  – Be flexible, be reasonable, be practical
Usability versus research?

• Approach similar to one used in classic research
• Different goals
• Usability testing is involved in building a successful product (resources, time, trade-offs)
  – Practical, large impact
  – Optimize interface after each iteration
• Usability study – Research study - Research on usability methods
A comparison

<table>
<thead>
<tr>
<th>Classical Research</th>
<th>Usability Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental design- isolate and understand specific phenomena, with the goal of generalization to other problems</td>
<td>Find and fix flaws in a specific interface, no goal of generalization</td>
</tr>
<tr>
<td>Experimental design- a larger number of participants is required</td>
<td>A small number of participants can be utilized</td>
</tr>
<tr>
<td>Ethnography- observe to understand the context of people, groups, and organizations</td>
<td>Observe to understand where in the interface users are having problems.</td>
</tr>
<tr>
<td>Ethnography- researcher participation is encouraged</td>
<td>Researcher participation is not encouraged in any way</td>
</tr>
</tbody>
</table>
### A comparison

<table>
<thead>
<tr>
<th>Classical Research</th>
<th>Usability Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethnography-longer-term research method</td>
<td>Short-term research method</td>
</tr>
<tr>
<td>Ethnography or experimental design—used to understand problems and/or answer research questions</td>
<td>Used in systems and interface development</td>
</tr>
<tr>
<td>Ethnography or experimental design—used at earlier stages, often separate (or only quasi-related) from the interface development process</td>
<td>Typically takes place in later stages, after interfaces (or prototype versions of interfaces) have already been developed</td>
</tr>
<tr>
<td>Used for understanding problems</td>
<td>Used for evaluating solutions</td>
</tr>
</tbody>
</table>
Research ABOUT usability testing

• So, usability testing IS different from traditional research
• But, there is also research ABOUT usability testing, with research questions such as:
  – Which methods are most effective
  – Which methods are most cost-effective
  – How many users are optimal
Usability engineering

Activities aiming to improve the ease of use of an interface

• Expert-based testing *(usability inspection)*
• Automated testing *(usability inspection)*
• User-based testing *(usability testing)*
Expert-based testing

- Structured inspections done by interface experts
- Before tests with users
- Confusing wording, inconsistent layout, obvious flaws
- Heuristic review
  - Compare interface with the rules
- Consistency inspections
  - Series of screens or web pages inspected
- Cognitive walkthrough
  - Experts perform the tasks (high-frequency and important/seldom)
- Guidelines review
  - Web Content Accessibility Guidelines
Heuristic evaluation

• Heuristic evaluation enables designers to evaluate an interface \textit{without} users
  – inspection, guided by a set of guidelines

• Economical technique to identify usability issues early in the design process
  – no implementation or users required
  – can be performed on existing interfaces
• Briefing
  – teach to evaluators; ensure each person receives same briefing.
  – become familiar with the UI and domain

• Evaluation period
  – compare UI against heuristics
  – spend 1-2 hours with interface; minimal 2 interface passes
  – take notes

• Debriefing session
  – Prioritize problems; rate severity
  – aggregate results
  – discuss outcomes with design/development team
1. Validity of system status
   - Are users kept informed about what is going on?
   - Is appropriate feedback provided within reasonable time about a user’s action?

2. Match between system and the real world
   - Is the language used at the interface simple?
   - Are the words, phrases and concepts used familiar to the user?

3. User control and freedom
   - Are there ways of allowing users to easily escape from places they unexpectedly find themselves in?

4. Consistency and standards
   - Are the ways of performing similar actions consistent?

5. Help users recognize, diagnose, and recover from errors
   - Are user messages helpful?
   - Do they use plain language to describe the nature of the problem and suggest a way of solving it?

6. Error prevention
   - Is it easy to make errors?
   - If so, where and why?

7. Recognition rather than recall
   - Are objects, actions and options always visible?

8. Flexibility and efficiency of use
   - Have accelerators (i.e. shortcuts) been provided that allow more experience users to carry out tasks more quickly?

9. Aesthetic and minimalist design?
   - Is any unnecessary and irrelevant information provided?

10. Help and documentation
     - Is help information provided that can be easily searched and easily followed?
• These heuristics are too general in some cases

• What about web pages? Ambient displays? Ubiquitous computing applications?

• Strong need for heuristics that are tailored to specific products.

• High quality content
• Often updated
• Minimal download time
• Ease of use
• Relevant to users’ needs
• Unique to the online medium
• Netcentric corporate culture

- Nielsen 1999 for commercial websites
Eight Golden Rules of Interface Design

Eight Golden Rules of Interface Design
As a result of Interface Design Studies, Ben Shneiderman proposed a collection of principles that are derived heuristically from experience and applicable in most interactive systems. These principles are common for user interface design, and as such also for web design.

1. Strive for consistency.
2. Enable frequent users to use shortcuts.
3. Offer informative feedback.
4. Design dialog to yield closure.
5. Offer simple error handling.
6. Permit easy reversal of actions.
7. Provide the sense of control. Support internal locus of control.
Pros and Cons of HE

• Pros
  – Very cost effective
  – Identifies many usability issues

• Cons
  – relies on interpretation of guidelines
  – guidelines may be too generic
  – needs more than one evaluator to be effective
Exercise

Groups of 5 students, 5-10 minutes

Go to the alarm app on your phone

– Prepare heuristic evaluation with 8 golden rules
– Conduct heuristic evaluation
Cognitive walkthrough involve simulating a user’s problem-solving process at each step in the human-computer dialog, checking to see if the user’s goals and memory for actions can be assumed to lead to the next correct action. —Nielsen and Mack, 1994
Why we use it

• Cognitive walkthrough enables a designer to evaluate an interface *without* users
  – a designer attempts to see the interface from the perspective of a user

• Low-investment technique to identify task-related usability issues early on
  – no implementation or users required
  – can be performed on existing interfaces
• **Identify task-related problems before implementation**
  – invest a little now, save a lot later

• **Enables rapid iteration early in design**
  – can do several evaluations of trouble points

• **Evaluations are only effective if your team**
  – has the right skill set
  – *wants to improve the design, not defend it*
Walkthrough Basics

• Imagine how well a user could perform tasks with your low-fidelity prototype

• Manipulate prototype as you go
  – evaluate choice-points in the interface
  – evaluate labels or options
  – evaluate likely user navigation errors

• Revise prototype and perform again
When to do the Walkthrough

• Have a low-fidelity prototype of the interface
• Know who the users are
• Have task descriptions
• Have scenarios designed to complete the task
  – you have a “functional” paper prototype
• Viable once the scenario and paper prototype are complete
Walkthrough Cons

• Is diagnostic, not prescriptive
• Focuses mostly on novice users
• Designers must put themselves in users mind
• Focus specifically on task-related issues
• The interactions are slower and not real
• Does not provide quantitative results

• A useful tool in conjunction with others
Automated usability testing

- Software that compare interface with the guidelines
- Produce report and/or fix the code
- Manual check often needed
- `<alt>` tag (alternative for graphics) but not if the text is appropriate
  - 'picture here'
- Number of fonts, avg. font size, deepest level of a menu
- Software applications: RAMP, InFocus
User-based testing

– Select representative users
– Select the setting
– Decide what tasks users should perform
– Decide what type of data to collect
– Before the test session (informed consent, etc.)
– During the test session
– Debriefing after the session
When to test

– Users perform tasks (early or later in the development)

– Formative testing
  • Low-level fidelity prototypes
  • How the user perceives an interface component?
  • Low-cost of paper prototypes, users comfortable to criticize

– Summative testing
  • Evaluate the effectiveness of specific design choices
  • High fidelity prototype

– Validation test
  • Before release, compare to benchmarks

– Formal approach
How many users?

– 5 users will find approximately 80% of problems
– 7 for small projects, 15 for medium-large projects
– Goal to find the major flows, that will cause the most problems and must be fixed
– How many users can we afford? How many users can we get? How many users do we have time for?
Location

- Lab
- Portable usability lab
- Remote studies (time, place)
Tasks

• Clear; no need for further explanation
• Tested
• One clear answer – way to do things
• Tasks that are performed often
• Critical tasks – logging
• No private, financial information
• Must be clear how to go to the next task
Measurement

– Task performance
– Time performance
– User satisfaction (validated survey tool)
– Average time to recover from an error
– Time spent using help
– Number of visits to the search feature
– Time spent on specific web page
– Typing speed
– Qualitative data
Testing session

• Preparation
• The session
• After the session

• Example – R2D2 project
Making sense of the data

– Write up the results and help influence the design of the specific interface
– Presentation to developers and managers
– Report should
  • Include all flaws
  • Priorities
  • For each flaw: describe the problem, present the data, priority, suggest a fix, estimate time/efforts for the fix
– Report structure
  • How you did usability testing and how you prepared
  • What happened during the testing
  • The implications and recommendations