Usability Evaluation
AISD 2005

What is usability?!!
- ISO usability standard 9241
  - Effectiveness - can you achieve what you want to?
  - Efficiency - can you do it without wasting effort?
  - Satisfaction - do you enjoy the process?
- Nielsen, Usability Engineering, 1993
  - Learnability - how easy is for a new user to use an interface and accomplish tasks?
  - Efficiency of use – how much time a user takes to perform his tasks, once he is familiar with the interface?
  - Memorability - how does the user use the system after a period of not using it?
  - Few and non catastrophic errors – how many errors occur, what is its severity and how easy is to recover from them?
  - Subjective satisfaction – what is the level of user satisfaction while interacting with the system?

Design life-cycle

Evaluation
- Accesses level in which design follows principles
- Motivates & supports (re)design process
- Should be present along conception, development and maintenance
- Different techniques to apply (depending on…)
  - Evaluation goals
  - Budget
  - Availability of final/real users
  - Evaluators expertise
  - State of development of the product
  - Laboratory or field studies
  - With or without tangible artifact

Cognitive walkthrough
- Proposed by Polson et al.
- Origin - code walkthrough
- Analyses actions user has to perform to complete task
- Implies description
  - Of system prototype
  - Of task to be performed by user
  - Of actions to perform in order to complete the task
  - Of system users, indicating their knowledge and experience

Cognitive walkthrough (2)
- For each actions evaluators ask:
  - Is the user trying to produce one of the possible effects of the action?
  - Is the user able to understand that the correct action is available?
  - Once the right action is identified, will the user notice that he/she is facing the correct action to produce the result he/she is trying to achieve?
  - Once the action is completed, will the user be able to recognize the system feedback?

<table>
<thead>
<tr>
<th>Evaluation stage</th>
<th>Resources</th>
<th>Experience</th>
<th>Conclusion</th>
<th>Time</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Poor</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Nice</td>
</tr>
<tr>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
</tbody>
</table>
Heuristic evaluation

- Developed by Jacob Nielsen and Rolf Molich
- Inspection method to critique the system
- Based in set of general and simple heuristics
  - Visibility of system status
  - Match between system and the real world
  - User control and freedom
  - Consistency and standards
  - Error prevention
  - Recognition rather than recall
  - Flexibility and efficiency of use
  - Aesthetic and minimalist design
  - Help users recognize, diagnose, and recover from errors
  - Help and documentation

Heuristic Evaluation (2)

- Proceedings...
  - Experts check interface individually, identifying potential problems
  - Conclusions from all evaluators are merged in a hierarchical set of usability problems
- Outcome - list of usability problems, with reference to
  - Violated principles
  - Severity
  - Frequency
  - Correction priority

Review-based evaluation

- Evaluation based on previous studies
- Results from the literature used to support or refute parts of design
- Care needed to ensure results are transferable to new design.

Model based evaluation

- Based in cognitive or design models
- GOMS - Goals, Operators, Methods and Selection
  - predicts user performance with a particular interface
- KLM - keystroke Level Model
  - facilitates the prediction of the time needed by users to conclude a certain task

User testing

- Most powerful method to perform design evaluation
- Provides empirical evidence/results about real tasks
- Selection of subjects, variables and hypothesis is vital
- Goal - demonstrate that initial suppositions are statistically confirmed and correct
- Development teams are more receptive to changes in the design if these are based in empirical tests

Talk aloud protocol

- Popular & simple way of gathering information about the way users interact with the system
- User observed performing his usual tasks, while evaluator, virtually invisible, records user actions
- Users are asked to
  - Perform a task
  - Externalize their thoughts
  - Describe what they are doing, what they think is happening and the reasons why they are taking their decisions
- Allows identification of important clues about the interface without requiring a great level of expertise
Retrospective testing

- Tries to answer the question why?
- Reflection about tasks performed
- Transcript or video played back for participant to comment
  - immediately => fresh in mind
  - delayed => evaluator has time to identify questions
- Useful to identify reasons for actions and alternatives

Quantification tools

- Applications installed on the server to
  - Collect and store data about user actions
- Log analysis allows descriptive statistics on
  - Frequency of system access
  - Name and size of downloaded or accessed files
  - Number of clicks
  - User paths
- Give clues about problematic areas but do not answer the question why

Interrogation techniques

- Obtains information about
  - User tasks and requisites
  - User interpretations of the system
  - User needs, preferences and experience
- Specially productive in detecting critical incidents
- Interviews
  - Data collected in a direct and structured way
  - Advantage - adjusting the dialogue to the context, interviewee, depth, interest and relevance to give to each topic
- Questionnaires
  - Less flexible than interviews
  - Questions, sometimes answers, defined a priori
  - Reaches large user group

Which to select, when and why??!!

Combination is important...

User testing

- Summative evaluation
- Average time needed to complete task
- Percentage of tasks concluded
- Number of errors
- Facts to validate system and justify changes

Majestic conclusions...

- Start evaluating as soon as possible
- Evaluate as often as possible
- Combine more than one technique
- Manage advantages and disadvantages of when used individually
- What to correct??!!
  - Not specific or easy...
  - Goals vs constraints => trade offs
  - Different people to please (programmers, designers, end users)
  - Deadlines
  - Money
Some bibliography