Usability Inspection Methods

CS160: User Interfaces
John Canny
Topics

• Discount Usability

• Cognitive Walkthroughs

• Heuristic Evaluation
Iterative Design

- Design
  - Brainstorming
  - Task analysis
  - Contextual inquiry

- Prototype
  - Low-fi, paper

- Evaluate
  - Low-fi testing,
  - Qualitative eval
  - Quantitative eval
Genres of assessment

**Automated**  Usability measures computed by software

**Empirical**  Usability assessed by testing with real users

**Formal**  Models and formulas to calculate measures

**Inspection**  Based on heuristics, skills, and experience of evaluators
Quantitative Testing is Costly

- User studies are very expensive – you need to schedule (and normally pay) many subjects.

- User studies may take many hours of the evaluation team’s time.

- A user test can easily cost $10k’s
“Discount Usability” Techniques

• Cheap
  – No special labs or equipment needed
  – The more careful you are, the better it gets

• Fast
  – On order of 1 day to apply
  – (Standard usability testing may take a week)

• Easy to use
  – Can be taught in 2-4 hours
“Discount Usability” Techniques

• Cognitive Walkthroughs
  – Put yourself in the shoes of a user
  – Like a code walkthrough

• Heuristic Evaluation
  – Assess interface based on a predetermined list of criteria

• Other, non-inspection techniques are on the rise
  – e.g., online remote experiments with Mechanical Turk
Topics

• Discount Usability

• Cognitive Walkthroughs

• Heuristic Evaluation
Cognitive Walkthrough

• Formalized technique for imagining user’s thoughts and actions when using an interface:

• “Cognitive walkthroughs 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, 1992)
Cognitive Walkthrough

• Given an interface prototype or specification, need:
  – A detailed task with a concrete goal, ideally motivated by a scenario
  – Action sequences for user to complete the task

• Ask the following questions for each step:
  – Will the users know what to do?
  – Will the user notice that the correct action is available?
  – Will the user interpret the application feedback correctly?

• Record: what would cause problems, and why?

From: Preece, Rogers, Sharp – Interaction Design
Cognitive Walkthrough Example

- **Task**: Find the call number and location of the latest edition of the book “Interaction Design” by Preece, Rogers & Sharp in the Berkeley library

- **Typical users**: Students who are familiar with the web, but not necessarily with the library or its website
Cognitive Walkthrough Example

- Step 1: Select library catalog.

- Will the user know what to do?

- Will user notice that action is available?

- Will user interpret feedback correctly?
Cognitive Walkthrough Example

- Step 2: Complete the search form
  - Will the user know what to do?
  - Will user notice that action is available?
  - Will user interpret feedback correctly?
Cognitive Walkthrough Example

- Step 3: Locate the right edition, click to detail screen

- Will the user know what to do?

- Will user notice that action is available?

- Will user interpret feedback correctly?
Cognitive Walkthrough Example

• Step 4: Locate call number and library location

• Will the user know what to do?

• Will user notice that action is available?

• Will user interpret feedback correctly?
Topics

• Discount Usability

• Cognitive Walkthroughs

• Heuristic Evaluation
Usability Heuristics

• “Rules of thumb” describing features of usable systems
  – Can be used as design principles
  – Can be used to evaluate a design

• Example: Minimize users’ memory load

• Pros and cons
  – Easy and inexpensive
    • Performed by experts
    • No users required
    • Catch many design flaws
  – More difficult than it seems
    • Not a simple checklist
    • Cannot assess how well the interface will address user goals
Heuristic Evaluation

- Developed by Jakob Nielsen (1994)

- Can be performed on working UI or on sketches

- Small set (3-5) of evaluators (experts) examine UI
  - Evaluators check compliance with usability heuristics
  - Different evaluators will find different problems
  - Evaluators only communicate afterwards to aggregate findings
  - Designers use violations to redesign/fix problems
Nielsen’s Ten Heuristics

- H2-1: Visibility of system status
- H2-2: Match system and real world
- H2-3: User control and freedom
- H2-4: Consistency and standards
- H2-5: Error prevention
- H2-6: Recognition rather than recall
- H2-7: Flexibility and efficiency of use
- H2-8: Aesthetic and minimalist design
- H2-9: Help users recognize, diagnose, recover from errors
- H2-10: Help and documentation
Where do they come from?
Where do they come from?

Figure 1. 3 Prominent Severity Scales

<table>
<thead>
<tr>
<th>Severity</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>I don't agree that this is a usability problem at all</td>
</tr>
<tr>
<td>1</td>
<td>Cosmetic problem only. Need not be fixed unless extra time is available</td>
</tr>
<tr>
<td>2</td>
<td>Minor usability problem: fixing this should be given low priority</td>
</tr>
<tr>
<td>3</td>
<td>Major usability problem: important to fix, so should be given high priority</td>
</tr>
<tr>
<td>4</td>
<td>Usability catastrophe: imperative to fix this before product can be released</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 1 problems prevent the completion of a task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2</td>
<td>Level 2 problems create significant delay and frustration</td>
</tr>
<tr>
<td>Level 3</td>
<td>Level 3 problems have a minor effect on usability</td>
</tr>
<tr>
<td>Level 4</td>
<td>Level 4 problems are more subtle, and point to a potential enhancement</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task failure</th>
<th>Prevents this user going further</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious problem</td>
<td>May hinder this user</td>
</tr>
<tr>
<td>Minor hindrance</td>
<td>Possible issue, but probably will not hinder this user</td>
</tr>
<tr>
<td>No problem</td>
<td>Satisfies the benchmark</td>
</tr>
</tbody>
</table>
Where do they come from?

• Start with a really big list of usability heuristics (101), then apply them to a whole lot of test interfaces.

• Use factor analysis (like clustering) to find strongest patterns, and eliminate redundant heuristics.
H2-1: Visibility of system status

- Keep users informed about what is going on. Example: response time
  - 0.1 sec: no special indicators needed
  - 1.0 sec: user tends to lose track of data
  - 10 sec: max. duration if user to stay focused on action
- Short delays: Hourglass
- Long delays: Use percent-done progress bars
  - (or CNN)
  - Overestimate usually better
H2-1: Visibility of system status

Users should always be aware of what is going on
- So that they can make informed decision
- Provide redundant information
H2-2: Match System & World

- Speak the users’ language
- Follow real world conventions
- Pay attention to metaphors

Bad example: Mac desktop
H2-3: User control & freedom

• Users don’t like to be trapped!

• Strategies
  – Cancel button (or Esc key) for dialog
  – Make the cancel button responsive!
  – Universal undo
H2-3: User control & freedom

• Offer “Exits” for mistaken choices, undo, redo
  – Don’t force the user down fixed paths

• Wizards
  – Must respond to Q before going to next step
  – Good for infrequent tasks (e.g., network setup) & beginners
  – Not good for common tasks (zip/unzip)
H2-4: Consistency and standards
H2-4: Consistency and Standards

NEW CUSTOMER

- **Give us your measurements**
  Take or ask someone to help take your measurements, by following our easy instructions. It takes just 5 minutes!

- **Send us your best fitting shirt** *(go directly to cart)*
  If you prefer not to take measurements, you can mail us your best fitting shirt. Our Master Tailor will take the necessary measurements and will return your shirt along with your order.

  *: Your shirt will be used for measurements only. We will not copy it.

- **Visit our NYC showroom** *(go directly to cart)*
  Contact us at contact@listerouge-paris.com to plan a private appointment at our New York showroom (Madison Ave & 40th St.).

EXISTING CUSTOMER

- **Your measurements are on file** *(go directly to cart)*
  If your last order fits perfectly, we will make the new shirts with exactly the same measurements.

- **If your measurements have changed**
  Simply note your measurements changes compared to your previous shirts.

http://www.useit.com/alertbox/application-mistakes.html
H2-5: Error Prevention

- Eliminate error-prone conditions or check for them and ask for confirmation.
H2-5: Error Prevention

- Aid users with specifying correct input
H2-5: Error Prevention

Don’t allow incorrect input

Lego Mindstorms

MIT Scratch
Preventing Errors

• Error types
  – Slips
    • User commits error during the execution of a correct plan.
      – Typos
      – Habitually answer “no” to a dialog box
      – Forget the mode the application is in
  – Mistakes
    • User correctly executes flawed mental plan
      – Usually the result of a flawed mental model – harder to guard against
H2-6: Recognition over Recall
H2-6: Recognition over Recall

- Minimize the user’s memory load by making objects, actions, and options visible.
H2-7: Flexibility and efficiency of use

http://www.iphoneuxreviews.com/?p=114
H2-8: Aesthetic and minimalist design

H2-8: Aesthetic and minimalist design

- No irrelevant information in dialogues
H2-8: Aesthetic and minimalist design

• Present information in natural order

![Diagram](https://via.placeholder.com/150)

- Occam’s razor
  - Remove or hide irrelevant or rarely needed information – They compete with important information on screen
    - Pro: Palm Pilot
    - Against: Dynamic menus
  - Use windows frugally
    - Avoid complex window management

From Cooper’s “About face 2.0”
H2-8: Aesthetic and minimalist design
<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Collation</th>
<th>Operator</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>blog_id</td>
<td>int(11)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_name</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_description</td>
<td>text</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_site_path</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_archive_path</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_archive_url</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_archive_type</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_days_on_index</td>
<td>int(11)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_language</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_file_extension</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_email_new_comments</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_email_new_pings</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_allow_comment_html</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_autodiscover_urls</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_sort_order_posts</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_sort_order_comments</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_allow_comments_default</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_allow_pings_default</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_server_offset</td>
<td>float</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_convert paras</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_convert paras</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_status_default</td>
<td>smallint(6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_allow_anon_comments</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_words_in_excerpt</td>
<td>smallint(6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_ping_weblogs</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_ping_blogs</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_ping_ones</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_blog_id</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_mt_update_key</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_autodiscover_links</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_welcome_msg</td>
<td>text</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_archive_tmpl_monthly</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_archive_tmpl_weekly</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_archive_tmpl_daily</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_archive_tmpl_individual</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_archive_tmpl_category</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_google_api_key</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_google_api_key</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_sanitize_spec</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_license</td>
<td>varchar(255)</td>
<td>latin1_general_ci</td>
<td>LIKE</td>
<td></td>
</tr>
<tr>
<td>blog_is_dynamic</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_require_comment_emails</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_allow_reg_comments</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_junk_score_threshold</td>
<td>float</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_moderate_pings</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_moderate_unreg_comments</td>
<td>tinyint(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blog_created_by</td>
<td>int(11)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
H2-9: Help users recognize, diagnose, and recover from errors
Good Error Messages

Please take note:

The file you are editing, "Prose.txt", which resides on volume "NetOne", is unavailable due to an unidentified network failure. No data has been lost, however.

Scope: The file will be automatically restored when the network connection is re-established.

Action: A local copy of "Prose.txt" will be saved to the volume titled "Drive C:" in folder "Local Save" on your Desktop. If you would like to save to a different location, press Save As... below; otherwise, just press OK.

More: This program does not have the ability to diagnose the problem further. Please contact your network administrator for further information.

From Cooper’s “About Face 2.0”
H2-9: Help users recognize, diagnose, and recover from errors
H2-10: Help and documentation

- Help should be:
  • Easy to search
  • Focused on the user’s task
  • List concrete steps to carry out
  • Not too long
Types of Help

• Tutorial and/or getting started manuals
  – Presents the system conceptual model
    • Basis for successful explorations
  – Provides on-line tours and demos
    • Demonstrates basic features

• Reference manuals
  – Designed with experts in mind

• Reminders
  – Short reference cards, keyboard templates, tooltips…
Types of Help

- Context sensitive help
- Search
Examples

Jaeger LeCoultre Reverso Grande Automatic Watch
Our Price: $6,249.99
List: $8,250.00 (24% off)
Review: ★★★★★

Invicta Men's Swiss Quartz Steel Watch
Our Price: $119.99
List: $475.00 (75% off)
Review: ★★★★★

Lucien Piccard Men's Chronograph White Dial Watch
Our Price: $129.99
List: $450.00 (71% off)
Review: ★★★★★

Stuhrling Alpine Men's Automatic Skeleton Watch
Our Price: $149.99
List: $795.00 (81% off)
Review: ★★★★★

Oris Williams F1 Team Chronograph Men's Watch
Our Price: $1,749.99
List: $2,575.00 (32% off)
Review: ★★★★★

Invicta Pro Diver Men's Automatic Steel Watch
Our Price: $99.99
List: $285.00 (65% off)
Review: ★★★★☆
Examples
Examples
Examples
Examples
Examples
The Process of Heuristic Evaluation
Phases of Heuristic Eval. (1-2)

• 1) Pre-evaluation training
  – Provide the evaluator with domain knowledge if needed

• 2) Evaluation
  – Individuals evaluate interface then aggregate results
    • Compare interface elements with heuristics
  – Work in 2 passes
    • First pass: get a feel for flow and scope
    • Second pass: focus on specific elements
  – Each evaluator produces list of problems
    • Explain why with reference to heuristic or other information
    • Be specific and list each problem separately
Phases of Heuristic Eval. (3-4)

• 3) Severity rating
  – Establishes a ranking between problems
    • Cosmetic, minor, major and catastrophic
  – First rate individually, then as a group

• 4) Debriefing
  – Discuss outcome with design team
  – Suggest potential solutions
  – Assess how hard things are to fix
Examples

• Typography uses mix of upper/lower case formats and fonts
  – Violates “Consistency and standards” (H2-4)
  – Slows users down
  – Fix: pick a single format for entire interface

  – Probably wouldn’t be found by user testing
Severity Rating

• Used to allocate resources to fix problems
• Estimates of need for more usability efforts
• Combination of Frequency, Impact and Persistence
• Should be calculated after all evaluations are in
• Should be done independently by all judges
Levels of 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
Severity Ratings Example

• 1. [H2-4 Consistency] [Severity 3]

• The interface used the string "Save" on the first screen for saving the user's file, but used the string "Write file" on the second screen. Users may be confused by this different terminology for the same function.
Debriefing

• Conduct with evaluators, observers, and development team members – start with general discussion

• Suggest improvements to address major usability problems

• Development team rates how hard things are to fix

• Make it a brainstorming session
  – Little criticism until end of session
Pros and Cons of Heuristic Evaluation
HE vs. User Testing

- HE is much faster
  - 1-2 hours each evaluator vs. days-weeks

- HE doesn’t require interpreting user’s actions

- User testing is far more accurate
  - Takes into account actual users and tasks
  - HE may miss problems & find “false positives”

- Good to alternate between HE & user-based testing
  - Find different problems
  - Don’t waste participants
Why Multiple Evaluators?

- Every evaluator doesn’t find every problem
- Good evaluators find both easy & hard ones
Number of Evaluators

• Single evaluator achieves poor results
  – Only finds 35% of usability problems
  – 5 evaluators find ~ 75% of usability problems
  – Why not more evaluators? 10? 20?
    • Adding evaluators costs more
    • Many evaluators won’t find many more problems

• But always depends on market for product:
  – popular products → high support cost for small bugs
Decreasing Returns

Problems Found

Benefits / Cost

Caveat: graphs are for one specific example!
Summary

• Heuristic evaluation is a discount method
• Have evaluators go through the UI twice
  – Ask them to see if it complies with heuristics
• Have evaluators independently rate severity
• Combine the findings from 3 to 5 evaluators
• Discuss problems with design team
• Cheaper alternative to user testing
  – Finds different problems, so good to alternate