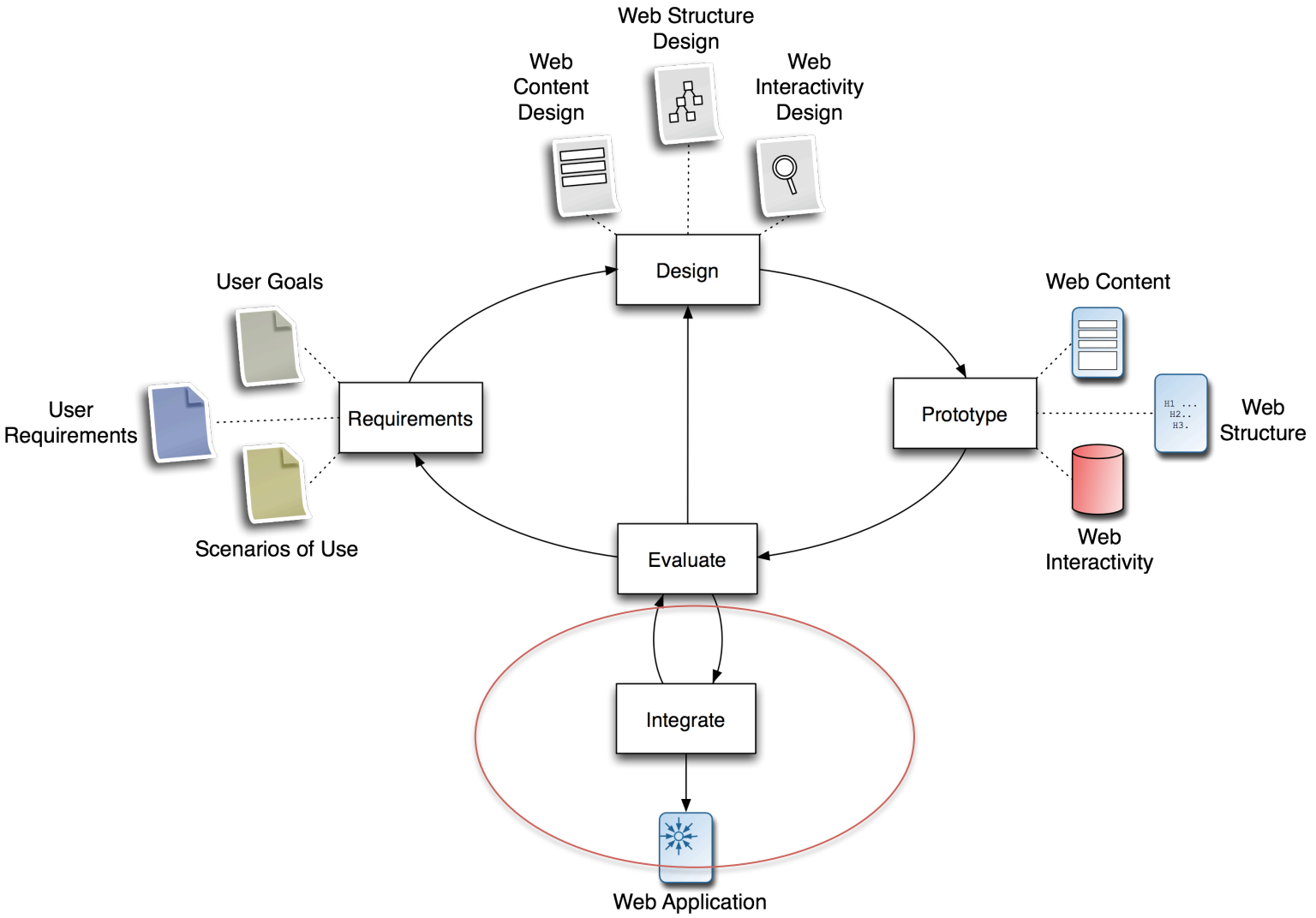
Expert Inspection Evaluations

HCI Research Group
Department of Computer Science

Remember that iterative cycle?



Evaluation

- After we have a design, it is important that we evaluate it
- Evaluation is the checking of the prototype for various usability properties
- We want to ensure that we are
  - Making a useful interactive system (meeting the functional requirements)
  - Making a usable interactive system (meeting usability requirements)

Types of Evaluation

- Expert Inspection Evaluation
  - Usability experts use a specific methodology to inspect the interface for various properties
  - Allow us to capture usability issues before putting it in front of users
- User Evaluation
  - We have users undertake tasks on an interface
  - Gold standard for evaluation allowing collection of high level tasks information or low level operation information

Expert Inspections

- Expert inspections can come in a variety of different forms:
  - Heuristic evaluations
  - Cognitive walkthroughs
  - Human error analysis
  - Human reliability analysis
  - Link analysis
- A host of other evaluation types are out there as well, but the first one is the most common and the one we will look at most closely in this module

Heuristic Evaluation

- Marketed originally as a “discount” usability method by its inventor Jakob Nielsen (you’ve heard that name before!)
- Purports to catch a substantial number of usability problems with 3-5 experts evaluating the interface
- Relies on rules of thumb that the experts evaluate the interface against
  - A lot of this work is done not by going through each heuristic, but instead by looking for problems in the interface and then backtracking to the heuristics

Heuristic Evaluation Process

- Briefing session
  - Experts are provided with a set of tasks, user journeys or at worst a set of screens to interact with
  - Often specified in collaboration with stakeholders and designers of application to ensure that priority is given to key tasks

Heuristic Evaluation (2)

- Evaluation Session
  - Each individual evaluator goes away on their own to work on the evaluation
  - Evaluation involves expert going through the interface
    - Usually begins with an overview inspection of each task
    - Then detailed inspection of each screen for issues
  - Expert records issues encountered, possible heuristics, and in some cases redesigns (not recommended)

Heuristic Evaluation (3)

- Debriefing Session
  - Evaluators come together with their lists of problems
  - Evaluations identify problems that are the same and reduce down to a set of unique problems
  - Evaluators rate each problem in its severity
    - Severity here is defined as a combination of the impact it will have on the users (will it annoy them or completely interrupt them?) as well as the frequency with which these will happen
    - Not entirely clear that all evaluators use the same weighting for these!

Heuristic Evaluation (4)

- Problem severity levels:
  - Cosmetic problem
    - An irritant to users which will come up occasionally
  - Minor problem
    - A problem which will interrupt users and their flow in the task; happens infrequently and users typically will know how to deal with it after first encounter
  - Major problem
    - A problem which will interrupt users substantially in completing their task and may take a substantial effort to solve or a problem which occurs frequently
  - Catastrophic problem
    - A show-stopper; a problem which keeps users from completing their tasks

Task Based User Evaluation Problem Summary

Problem Description	Severity	Solution
User can't understand if their pictures are secure	Catastrophe	Identify ways to change feedback so users know tasks are completed
Overwhelming number of items in privacy settings	Major	Categorization of links into appropriate groupings Aggregate some settings
Privacy settings hidden under account menu not in profile area	Major	Make link under profile more obvious to the user
Privacy settings under table labelled "notifications" - very strange labelling	Minor	Change label to Notifications and Privacy
Page "My profile" not the same as the menu item selected - disorienting	Cosmetic	Change menu label

Nielsen's Heuristics

- Visibility of system status
  - Are users kept informed about what is going on?
  - Is appropriate feedback provided within reasonable time about a user's action?
- Match between system and the real world
  - Is the language used in the system simple?
  - Are the words, phrases and concepts used familiar to the user?
- User control and freedom
  - Are there ways of allowing users to easily escape from places they unexpectedly find themselves in?
- Consistency and standards
  - Are the ways of performing similar actions consistent?

Nielsen's Heuristics (2)

- Help users recognize, diagnose and recover from errors
  - Are error messages helpful?
  - Do they use plain language to describe the nature of the problem and suggest a way of solving it?
- Error prevention
  - Is it easy to make errors?
  - If so, where and why?
- Recognition rather than recall
  - Have shortcuts been provided that allow more experienced users to carry out tasks more quickly?
- Aesthetic and minimalist design
  - Is any unnecessary and irrelevant information provided?
- Help and documentation
  - Is help information provided that can be easily searched and easily followed?

Improvements: Heuristics

- Nielsen's heuristics are criticised for being too general, or too impractical to use across all systems
  - There are dozens of them out there on the web
- Some are very specialised, where they have been created to try to address a particular domain:
  - Games, mobile, safety critical systems, medical devices, multimedia …
- However, very few have been validated – worse many are just rewording of Nielsen's work

Heuristics for Interactive Websites

- At the University of York, we have created our own set of heuristics based on user data from interactive websites
- Analysed 935 usability problems collected from 30 users on interactive websites
- Sorted the problems into similar issues that kept recurring
- Created an overall structure and new set of heuristics that fall into the following categories:
  - Physical Presentation
  - Content
  - Information Architecture
  - Interactivity
- Highly refined set of heuristics specifically for websites
- We will use these in practical on Friday

Improvements: Method

- Remember when I said that this was going to be a discount method?
- Consider how much time it takes:
  - Briefing: 5 people x 1 hour
  - Evaluation: 5 people x 2 hours
  - Debriefing: 5 people x 2 hours
- So we have a “discount” method that is taking approximately 25 person hours for 1 website – that seems like a long time still
- What improvements are there that we can have in the method?

Improvement: Method (2)

- What if we were to find a way to eliminate the need for the debriefing?
- We would need to achieve the following:
  - Get all of the problems matched together
  - Get all of the problems rated in a consistent fashion
  - Allow all evaluators to contribute to problem pool
  - Avoid social loafing issues
  - Avoid influencing ratings

Collaborative Heuristic Evaluation

- Changing the evaluation to being collaborative in a group can improve the method (and make it more fun)
- Evaluation session:
  - All evaluators work their way through the website together
  - 1 person drives the application
  - All members can suggest problems, preferably in a neutral way, and avoiding redesign
  - 1 person records the problems
  - All members privately rate the problem on their own sheet using a scale from 0-4, where 0 means “not a problem
- At the end of the evaluation you have a complete set of problems already collected and a mean score from across the evaluators

Strengths of Heuristic Evaluation

- Usually picks up the majority of the major issues – keeps users from encountering obvious problems in design
- Catches a large number of problems relatively quickly Many experts means that many different perspectives on one application can be recorded
- Severity lists create a priority list for fixing problems

Weaknesses of Heuristic Evaluation

- Guaranteed to not pick up all of the problems
- Can be a danger of not using heuristics and just stating opinions
- Issues around specifying redesign instead of recording the problems
- Issues in calibration between experts regarding what is a cosmetic vs. minor vs. major vs. catastrophic – resolved through practice and discussion of standards

Readings

- Nielsen Norman Group Website on Heuristic Evaluation
  - How to Conduct a Heuristic Evaluation - http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/
  - 10 Usability Heuristics for User Interface Design - http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/
- Nielsen Publications
  - Nielsen, J., and Molich, R. (1990). Heuristic evaluation of user interfaces, Proc. ACM CHI'90 Conf. (Seattle, WA, 1-5 April), 249-256.
  - Nielsen, J. 1992. Finding usability problems through heuristic evaluation. Proceedings ACM CHI'92 Conference (Monterey, CA, May 3-7), 373-380.
  - Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J., and Mack, R.L. (Eds.), Usability Inspection Methods. John Wiley & Sons, New York, NY.

Readings (2)

- Heuristics for interactive websites
  - Petrie, H., & Power, C. (2012, May). What do users really care about?: a comparison of usability problems found by users and experts on highly interactive websites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2107-2116). ACM.
- Collaborative Heuristic Evaluation
  - Petrie, H., & Buykx, L. (2010). Collaborative Heuristic Evaluation: improving the effectiveness of heuristic evaluation. In Proceedings of UPA 2010 International Conference. Omnipress. Available at: http://upa. omnibooksonline. com/index. htm.