

Evaluation Plans

Chapters 10 and 11
Interaction Design

Overview

- Intro to evaluation
- Case Studies
- Evaluation Techniques

Two main types of evaluation

- **Formative evaluation** is done at different stages of development to check that the product meets users' needs.
- **Summative evaluation** assesses the quality of a finished product.

Our focus is on formative evaluation

What to evaluate

Iterative design & evaluation is a continuous process that examines:

- Early ideas for conceptual model
- Early prototypes of the new system
- Later, more complete prototypes

Designers need to check that they understand users' requirements.

When to evaluate

- Throughout design
- From the first descriptions, sketches etc. of users needs through to the final product
- Design proceeds through iterative cycles of 'design-test-redesign'
- Evaluation is a key ingredient for a successful design.

Case Studies

- OMS (Olympic Messaging System)
- HutchWorld
- Engineering Representational System
 - VesselWorld; GrewpTool; Cedar

Evaluating the 1984 OMS

- Early tests of printed scenarios & user guides
- Early simulations of telephone keypad
- An Olympian joined team to provide feedback
- Interviews & demos with Olympians outside US
- Overseas interface tests with friends and family.
- Free coffee and donut tests
- Usability tests with 100 participants.
- A 'try to destroy it' test
- Pre-Olympic field-test at an international event
- Reliability of the system with heavy traffic

Hutchworld

- Enables cancer patients, their caregivers, family, and friends to chat with one another
- tell their stories
- discuss their experiences and coping strategies
- Gain emotion and practical support
- Developed by Microsoft's Virtual Worlds Research group and librarians and clinicians at The Fred Hutchinson Cancer Research Center in Seattle, Washington

Early forms of data gathering

- Learn about patient experience
- Interviewed potential users
 - Patients, caregivers, family, friends, clinicians, and social support staff
- Also observed daily activity in clinic and hospital
- Read research literature, talked to experts, and former patients, ...

Some initial ideas

- Hutchworld should be available any time of day or night regardless of geographical location
- Virtual communities
 - Participants more open and uninhibited
 - Potential for misunderstanding is higher
- But research showed, for example, women with breast cancer who received group therapy lived on average twice as long as those who did not

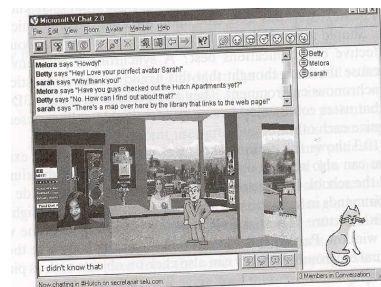
- Avatars
- List of commands
- List of participants
- Textual chat
- Participants can move their avatars and make them gesture to tour the virtual environment
- Also can click on objects to interact with them

Early Prototype



Second prototype

- Only lobby fully developed

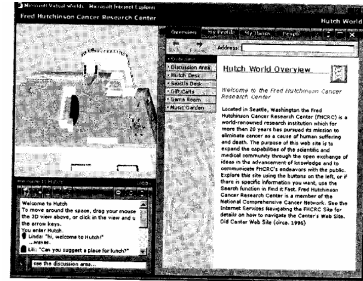


Test 1

- Early observations onsite
 - 6 computers set up
 - Simple scaled-back prototype of HutchWorld build using existing product, Microsoft V-Chat
 - Team observed the general usage of prototype
- What was learned?
 - No critical mass
 - Many patients didn't want simultaneous chatting
 - Computers also used to play games and search web for cancer sites
 - More unified site needed

Re-Design

- Support more asynchronous communication
- Second version functioned more as a portal to info-retrieval tools and communication tools, games, and other types of entertainment
- Also incorporated bulletin board, text-chat, and web page creation tool



Development of HutchWorld

- Many informal meetings with patients, carers & medical staff early in design
- Early prototype was informally tested on site
- Designers learned a lot e.g.
 - language of designers & users was different
 - asynchronous communication was also needed
- Redesigned to produce the portal version

Usability Tests

- Ran usability test in Microsoft usability labs
- 7 participants: 4 male, 3 female
- Subjects worked independently and provided running commentary
 - Commentary recorded on video and so were screens
- Microsoft evaluator watch through one-way mirror
 - Participants and evaluator interacted via microphone and speakers

Usability testing

- 5-minute exploration period then subjects asked to complete a series of *structured tasks*
 - How users' identity was represented
 - Communication
 - information searching
 - entertainment
- User satisfaction questionnaire
 - What did you like about HutchWorld?
 - What did you not like about HutchWorld?
 - What did you find confusing or difficult to use in HutchWorld?
 - How would you suggest improving HutchWorld?
- Triangulation to get different perspectives

Task #1: Explore HutchWorld

Your first task is to spend five minutes exploring HutchWorld.

A. First, open HutchWorld.

B. Now, explore!

Remember, tell us what you are looking at and what you are thinking about as you are exploring HutchWorld.

Task #2: All about Your Identity in HutchWorld

A. Point to the 3 dimensional (3D) view of HutchWorld.

B. Point at yourself in the 3D view of HutchWorld.

C. Get a map view in the 3D view of HutchWorld.

D. Walk around in the 3D view: go forward, turn left and turn right.

E. Change the color of your shirt.

F. Change some information about yourself, such as where you are from.

Task #3: All about Communicating with Others

- Send someone an email.
- Read a message on the HutchWorld Bulletin Board.
- Post a message on the HutchWorld Bulletin Board.
- Check to see who is currently in HutchWorld.
- Find out where the other person in HutchWorld is from.
- Make the other person in HutchWorld a friend.
- Chat with the other person in HutchWorld.
- Wave to the other person in HutchWorld.
- Whisper to the other person in HutchWorld.

Task #4: All about Getting Information

- Imagine you have never been to Seattle before. Your task is to find something to do.
- Find out how to get to the Fred Hutchinson Cancer Research Center.
- Go to your favorite website. (Or go to Yahoo: www.yahoo.com)
- Once you have found a website, resize the screen so you can see the whole web page.

Findings from the usability test

- The back button didn't always work
- Users didn't pay attention to navigation buttons
- Users expected all objects in the 3-D view to be clickable.
- Users did not realize that there could be others in the 3-D world with whom to chat,
- Users tried to chat to the participant list.

Participant number:	1	2	3	4	5	6	7	Average
Background Information								
Sex	F	F	M	M	F	M	M	3F, 4M
Age	37	41	43	54	46	44	21	40.9
years of college	4	2	4	4	4	1	2	3.0
hours of chat use in past year	0	3	0	0	365	200	170	105.4
hours of web use in past year	9	11	36	208	391	571	771	285.3
Structured Tasks								
Identify 3D view	1	1	1	1	1	1	1	1.0
Identify self in 3D view	1	2	1	1	1	1	1	1.1
Get a map view of 3D view	1	2	2	1	2	3	1	1.7
Walk in 3D view	1	3	2	1	3	2	1	1.9
Change color of shirt	1	1	3	3	2	3	2	2.1
Change where self is from	1	1	3	1	1	3	1	1.6
Find place to send email	1	3	3	1	3	2	2	2.1
Read a bulletin board message	2	1	3	1	1	1	-	1.5
Post a bulletin board message	1	3	3	3	2	2	-	2.3
Check to see who is currently on	1	3	1	3	2	3	2	2.1
Find out where the other person is from	1	1	2	1	1	3	2	1.6
Make the other person a friend	1	1	3	1	1	2	1	1.4
Chat with the other person	3	1	3	1	1	3	1	1.9
Wave to the other person	1	1	1	1	1	1	1	1.0
Whisper to the other person	1	3	2	2	1	2	1	1.7
Find something to do in Seattle	2	1	2	1	1	1	2	1.4
Find out how to get to FHCRC	1	3	3	2	1	1	2	1.9
Go to a website	1	3	2	3	3	1	1	2.0
Resize web screen	1	3	2	2	2	3	1	2.0
Find a game to play	1	1	2	1	1	1	2	1.3
Send self a gift	1	3	3	3	3	3	3	2.7
Open gift	3	1	2	3	3	3	3	2.6

1 -- easy
 2 -- ok
 3 -- difficult
bold -- needed help

Issue#	Issue Priority	Issue	Recommendation
1	high	Back button sometimes not working.	Fix back button.
2	high	People are not paying attention to navigation buttons.	Make navigation buttons more prominent.
3	low	Font too small, hard to read for some people.	Make it possible to change font size. Make the font colors more distinct from the background color.
4	low	When navigating, people were not aware overview button would take them back to the main page.	Change the overview button to home button, change the words of the overview page accordingly.
5	medium	"Virtual world" wording in login screen confusing.	Change wording to "HutchWorld".
6	high	People frequently clicking on objects in 3D view expecting something to happen.	Make the 3D view have links to web pages. For example, when people click on the help desk in 3D view, area should show the 3D link information.
7	low	People do not readily find map view button.	Make the icon on the map view button more map-like.
8	medium	Moving avatar with mouse took some getting used to.	Encourage the use of the keyboard. Mention clicking and dragging the avatar in the welcome.
9	low	People wanted to turn around in 3D view, but it was awkward to do so.	Make one of the chat buttons a button that lets you turn around.
10	medium	Confusion about the real world/virtual world distinction.	Change wording of overview description, to make clear HutchWorld is a "virtual" place made by people. Use FHCRC, and a place where anybody can go.
11	high	People do not initially recognize that other real people could be in HutchWorld, that they can talk to them and see them.	Change wording of overview description, to make clear HutchWorld is a place to "chat" with others who are "present" in the virtual HutchWorld.
12	high	People not seeing/finding the chat window. Trying to chat to people from the people list where other chat-like features are (whisper, etc.)	Make chat window more prominent. Show how link chat. Use feedback of navigation for chat window. Change wording of chat window. Change wording of chat window. Instead of type to chat here, speak here, type to chat here.

Key points

- Evaluation & design are closely integrated in user-centered design.
- Some of the same techniques are used in evaluation & requirements but they are used differently (e.g., interviews & questionnaires)
- Triangulation involves using a combination of techniques to gain different perspectives
- Dealing with constraints is an important skill for evaluators to develop.

Evaluation Techniques

- GOMS
- Cognitive Walkthrough
- Questionnaires
- Interviews
- Ethnography
 - Video Taping; Transcript & Replay

Evaluation Techniques

Observing Users	Asking users	Asking Experts	User Testing	Modeling users' task performance
Transcript & Replay	Interviews	Inspection: Shneiderman's 8 Golden Rules Nielsen's 10 Design Principles Guidelines for web page layout	Testing typical users doing typical tasks in laboratory setting	GOMS Fahrenheit ↔ Celsius
Video Taping	Questionnaires	Cognitive Walkthrough	Try to destroy it sessions	
Users talk aloud as they use interface				

Representational System

1. A set of representational media available to the participants.
2. A set of internal or external, private or shared, representations
3. A set of procedures for communicating, recording, modifying, transcribing, and aligning multiple, partial representations of the shared context.

Classroom

1. Chalkboard, books, student notebooks, laptops,
2. What is on the chalkboard versus what is in the notebook
3. Students take notes; power point slides are posted on class website

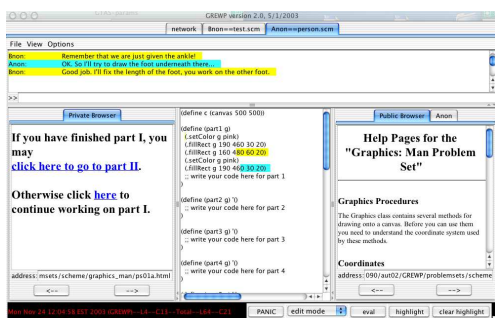
Basic Methodology (For Re-engineering the Rep. Sys.)

- Online practice is grounded in the representational system provided by a groupware system.
- Transcripts are collected of online user behavior.
- Identify weak spots in the representational system
 - Coordination work & cognitive load
- Re-engineer the representational system
- Initially applied to VesselWorld
 - Work done with Landsman, Feinman, Introne

Engineering Representational System (Evaluation / Development Plan)

- Requirements gathering
- Iteratively build prototype
 - Simple & generic, but provides replay
 - Read literature
 - Group design evaluation sessions
 - Inspection & Cognitive Walkthrough
 - Pounding within group; pounding by outsiders
- Pilot Study to collect transcript data
- Analysis & Re-Design of Representational System

GrewpTool



Iteratively Designing the Prototype

- In previous HCI class two groups of students had done term project for TA's to tutor students online
- VesselWorld, replay
- Interest in collaborative learning
- Initial designs the interaction between students were more structured
 - Read through literature on collaborative editing
 - Why? Mine for good design ideas to start with

Example of a collaborative editor

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

Initial Version of GrewpTool

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

Pilot study evaluation

- 6 students used GHT in pairs
 - Place in individual terminals out of each other's sight
 - Two sessions per pair; each lasting two hours
 - Session 1: Code webpage using HTML
 - Session 2: Simple application using Jscheme
- We were able to replay all the sessions

Evaluation

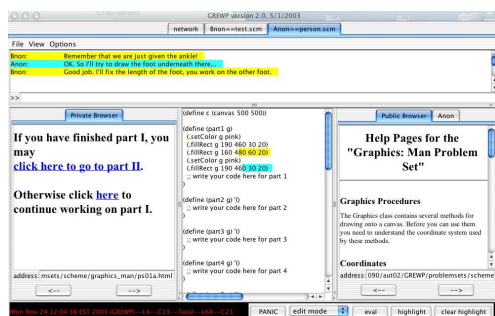
Issues

- Co-browsing was hard; typed URLs into chat window
- Whiteboard never used
- Students wanted to be able to more easily see what their partners were up doing.
- Needed to be able to capture the attention of their partners

Design Changes

- Watch versus edit mode
- Co-Browsing Tabs
- Removed whiteboard
- Added panic button

GrewpTool

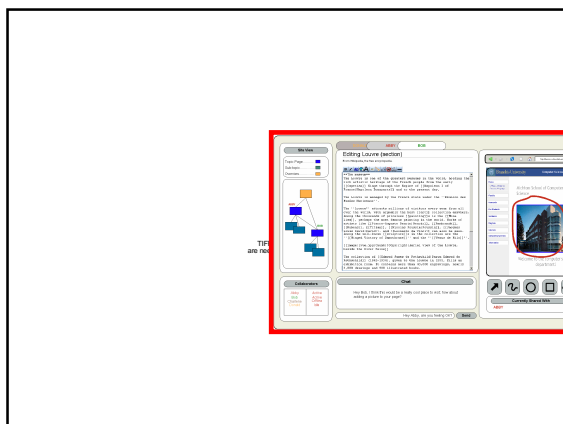
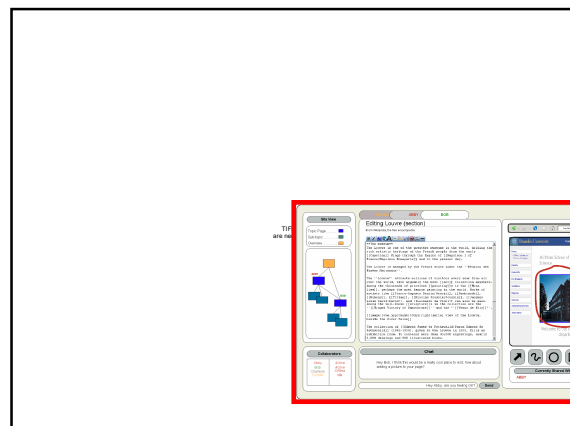
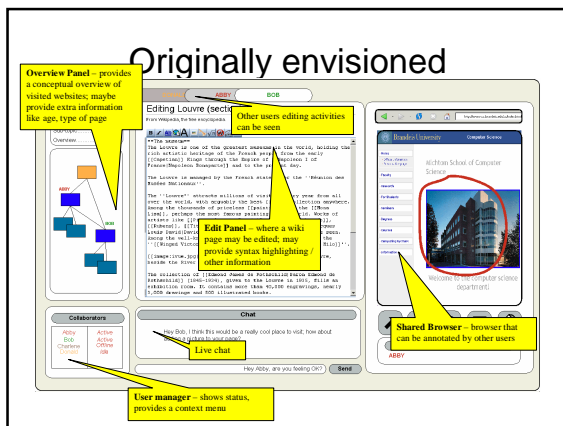
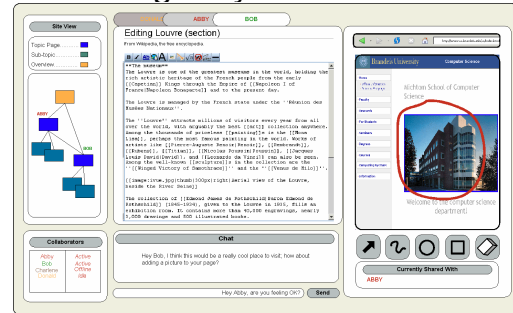


Development & Evaluation Plan for Cedar

Cedar

- A platform for studying online collaboration
 - Both same time / different place & different time / different place
 - Support code writing, website construction
 - An application wrapper around a Wiki web, that provides additional collaborative tools (e.g., Wikipedia)
 - Use Thyme & Sage toolkits to construct
- Also use in classroom
 - Computational Cognitive Science (data)
 - Internet & Society (website construction)
 - COSI 11: Intro to java coding
- With Johann Larusson, Josh Introne

Originally envisioned



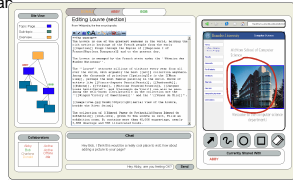
Discussion



QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

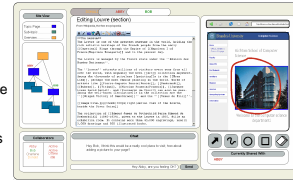
Computational Cognitive Science (COSI 111)

- Teams of students use Cedar as shell to develop an application (e.g., trip planner)
 - Collect replayable data of subjects using application
 - Last time, same time / different place
- Analysis of data
 - Last time emphasize coordination issues and referential structure of discourse
 - Re-Design



Internet & Society (COSI 33b)

- Construct personal homepage to play with notions of online identity
 - Not necessarily for yourself
- Term project (teams): Develop website on some topic
 - Create list of related website and evaluate them in terms of content & design
 - Each member of team does a term paper on some part of their topic
 - Term papers are organized as part of website



Research Issue: Rebuilding Shared Context

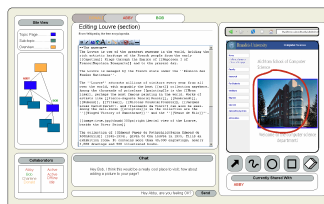
- Collaboration on longer tasks
- Asynchronous/synchronous
 - Need to integrate separate work
 - Must rebuild context for each synchronous collaboration period
- How can we better facilitate this (for software developers)?
- Initially developed with Mike Head

Context Integration

- Merging the work done separately
- Understanding of completed work so far
- Understanding of the assigned task
- Collaborators plan for future work
- ...

Context integration as paired programming

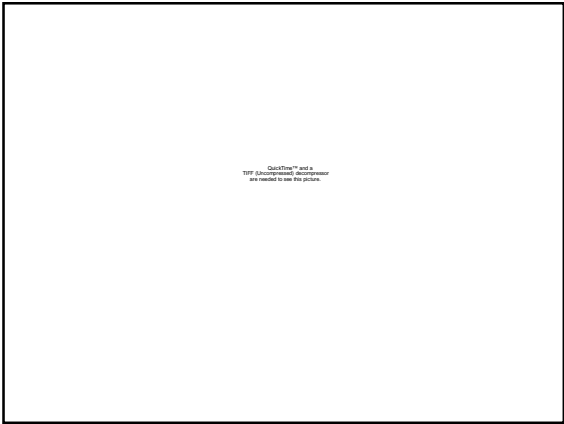
- Distributed Pair Programming
- Planning is like programming [merging in particular]
- Two (possibly more) programmers
- Working on the same file
- Synchronously/Asynchronously
- Remotely



Experiment

- Two programmers
 - Work remotely on an assigned task in three phases
 - 1) Synchronous design and analysis
 - Reading the problem
 - dividing up the work
 - 2) Asynchronous work
 - Coding separately
 - 3) Synchronous integration
 - Pull together the separate pieces of work

CedarTime™ and a 100% CD-ROM environment are needed to see this picture.



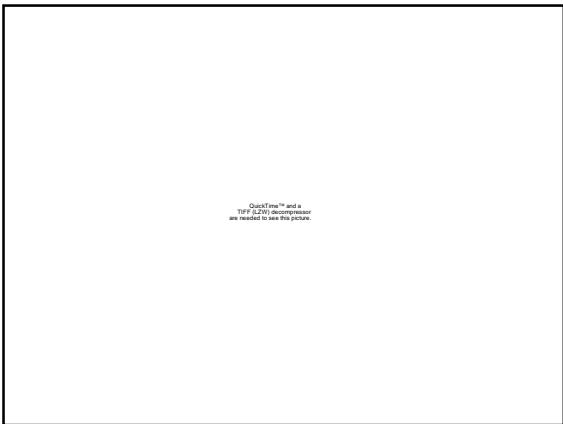
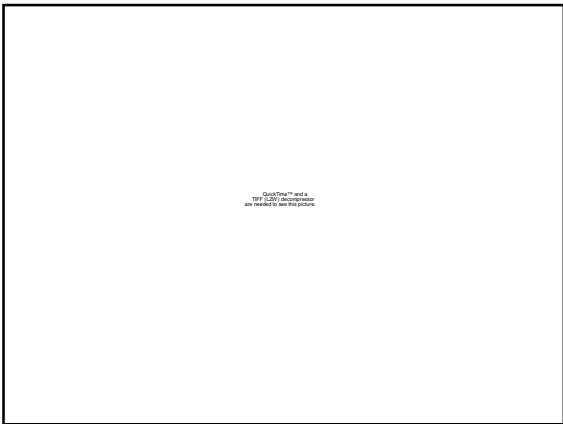
Evaluation/ Re-Design Plan???

- Evaluation so far:
 - Inspection
 - Group hack sessions
 - COSI 125 survey critique
 - Walkthrough with walk-bys
- Is Cedar within edit distance of sample class projects and research tasks?
 - Design representative task(s) for evaluation
- By early May: Pounding session
 - Make sure replay works
 - ID major problems
 - What else?
- To be continued

Observing Users	Asking users	Asking Experts	User Testing	Modeling users' task performance
Transcript & Replay	Interviews	Inspection: Shneiderman's 8 Golden Rules Nielsen's 10 Design Principles Guidelines for web page layout	Testing typical users doing typical tasks in laboratory setting	GCMS Fahrenheit ↔ Celsius
Video Taping	Questionnaires	Cognitive Walkthrough	Try to destroy it sessions	
Users talk aloud as they use interface				

Test Interface

- Want replay
- Design task to test various features
 - Both asynchronous & synchronous
- Two tests of interface???



Task for users???

- Update article by adding info on ...
- Re-design webpage using guidelines
- Both synchronous & asynchronous

© 2000-2001 Microsoft Corporation. All rights reserved. Microsoft, MSN, and MSN Messenger are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

Tasks

- Edit file
- View webpage you are editing in browser
- Save changes
- Chat
- Look at a page your partner is editing
- Navigate in browser
-