Evaluation Techniques

Overview

Evaluation

- tests usability and functionality of system
- occurs in laboratory, field and/or in collaboration with users
- evaluates both design and implementation

Evaluation should be considered at all stages in the design life cycle.

Goals of Evaluation

- assess extent of system functionality
- assess effect of interface on user
- identify specific problems

Styles of Evaluation

Laboratory studies

Advantages:

- specialist equipment available
- uninterrupted environment

Disadvantages:

- lack of context
- difficult to observe several users cooperating

Appropriate

- if system location is dangerous or impractical
- for constrained single user systems
- to allow controlled manipulation of use

Field Studies

Advantages:

- natural environment
- context retained (though observation may alter it)
- longitudinal studies possible

Disadvantages:

- distractions
- noise

Appropriate

- where context is crucial
- for longitudinal studies

Participatory Design

User is an active member of the design team.

Characteristics

- context and work oriented rather than system oriented
- collaborative
- iterative

Methods

- brain storming
- storyboarding
- workshops
- pencil and paper exercises

Cognitive Walkthrough

Proposed by Polson et al.

- evaluates design on how well it supports user in learning task
- usually performed by expert in cognitive psychology
- expert 'walks though' design to identify potential problems using psychological principles
- forms used to guide analysis

Cognitive Walkthrough (cont.)

For each task walkthrough considers

- what impact will interaction have on user?
- what cognitive processes are required?
- what learning problems may occur?

Analysis focuses on goals and knowledge: does the design lead the user to generate the correct goals?

An example is expanded in Section 11.4.1.

Heuristic Evaluation

Proposed by Nielsen and Molich.

- usability criteria (heuristics) are identified
- design examined by experts to see if these are violated

Example heuristics

- system behaviour is predictable
- system behaviour is consistent
- feedback is provided

Heuristic evaluation 'debugs' design.

Review-based evaluation

- Results reported in the literature are used to support or refute parts of design.
- Care is needed to ensure results are transferable to new design.

Model-based evaluation

- Cognitive models are used to filter design options. E.g. GOMS prediction of user performance.
- Design rationale can also provide useful information in evaluating designs.

Evaluating Implementations

Requires an artefact — simulation, prototype, full implementation.

Experimental evaluation

- controlled evaluation of specific aspects of interactive behaviour
- evaluator chooses hypothesis to be tested
- a number of experimental conditions are considered which differ only in the value of some controlled variable
- changes in behavioural measure are attributed to different conditions

Experimental factors

Subjects

- representative
- sufficient sample

Variables

- independent variable (IV) — characteristic changed to produce different conditions. E.g. interface style, number of menu items.
- dependent variable (DV) — characteristics measured in the experiment. E.g. time taken, number of errors.

Experimental factors (cont.)

Hypothesis

- prediction of outcome framed in terms of IV and DV
- null hypothesis: states no difference between conditions — aim is to disprove this

Experimental design

- within groups design — each subject performs experiment under each condition. Transfer of learning possible but less costly and less likely to suffer from user variation.
- between groups design — each subject performs under only one condition. No transfer of learning but more users required and variation can bias results.

Analysis of data

- look at data
- save original data

Choice of statistical technique depends on

- type of data
- information required

Type of data

- discrete — finite number of values
- continuous — any value

Analysis of data (cont.)

Types of test

- parametric
  - assume normal distribution
  - robust
  - powerful
- non-parametric
  - do not assume normal distribution
  - less powerful
  - more reliable
- contingency table
  - classify data by discrete attributes and count number of data items in each group

Analysis of data (cont.)

What information is required?

- is there a difference?
- how big is the difference?
- how accurate is the estimate?

Parametric and non-parametric tests address mainly first of these.

Worked examples of data analysis are given in Section 11.5.1.

Table 11.1 summarizes main tests and when they are used.

Observational Methods

Think Aloud

- user observed performing task
- user asked to describe what he is doing and why, what he thinks is happening etc.

Advantages

- simplicity —requires little expertise
- can provide useful insight
- can show how system is actually used

Disadvantages

- subjective
- selective
- act of describing may alter task performance

Observational Methods (cont.)

Cooperative evaluation — variation on think aloud

- user collaborates in evaluation
- both user and evaluator can ask each other questions throughout

Additional advantages

- less constrained and easier to use
- user is encouraged to criticize system
- clarification possible

Observational Methods (cont.)

Protocol analysis methods

- paper and pencil
  - cheap
  - limited to writing speed
- audio
  - good for think aloud
  - difficult to match with other protocols
- video
  - accurate and realistic
  - needs special equipment
  - obtrusive
- computer logging
  - automatic and unobtrusive
  - large amounts of data difficult to analyze

Observational Methods (cont.)

- user notebooks
  - coarse level and subjective
  - useful insights
  - good for longitudinal studies

Mixed use in practice.

Transcription of audio and video difficult and requires skill.

Some automatic support tools available

- EVA
- Workplace project

Observational Methods (cont.)

Post task walkthrough

- user reflects on action after the event
- used to fill in intention

Advantages

- analyst has time to focus on relevant incidents
- avoid excessive interruption of task

Disadvantages

- lack of freshness
- may be post-hoc interpretation of events

Query Techniques

- informal and subjective
- cheap

Interviews

Analyst questions user on one to one basis, usually based on prepared questions.

Advantages

- can be varied to suit context
- issues can be explored more fully
- can elicit user views and identify unanticipated problems

Disadvantages

- very subjective
- time consuming

Query Techniques (cont.)

Questionnaires

Set of fixed questions given to users.

Advantages

- quick and reaches large user group
- can be analyzed more rigorously

Disadvantages

- less flexible
- less probing

Need careful design

- what information is required?
- how are answers to be analyzed?

Questionnaires (cont.)

Styles of question

- general
- open-ended
- scalar
- multi-choice
- ranked

Choosing an Evaluation Method

Factors to consider (see also Tables 11.3-11.5)

- when in cycle is evaluation carried out? design vs implementation
- what style of evaluation is required? laboratory vs field
- how objective should the technique be? subjective vs objective
- what type of measures are required? qualitative vs quantitative
- what level of information is required? high level vs low level
- what level of interference? obtrusive vs unobtrusive
- what resources are available? time, subjects, equipment, expertise

Tables 11.3–11.5 rate each techniques along these criteria.