BINUS UNIVERSITY

Course : COMP6176 / Human - Computer Interaction

Year : 2019

ADVANCED EVALUATION

SESSION 12

People
Innovation
Excellence

LEARNING OUTCOMES

- LO 5: Evaluate the user interfaces of interactive software

People
Innovation
Excellence

BINUS UNIVERSITY

People
Innovation
Excellence

OUTLINE

- Introduction
- Usability Testing
- Conducting Experiments
- Field Studies
- Inspections
- Analytics
- Predictive Models

BINUS UNIVERSITY

People
Innovation
Excellence

INTRODUCTION

- The evaluation studies that take place in a spectrum of settings, from controlled laboratories to natural settings.
- We focus on **usability testing** which takes place in usability labs, **experiments** which take place in research labs, and **field studies** which take place natural settings such as people's homes, work, and leisure environment.
- And we introduce methods that are based on understanding users through knowledge codified in heuristics, or data collected remotely, or models that predict users' performance.

BINUS UNIVERSITY

People
Innovation
Excellence

USABILITY TESTING

- The usability of products has traditionally been tested in controlled laboratory settings.
- It has been commonly used to evaluate desktop applications, such as websites, word processors, and search tools.
- As mentioned in Session 11, a combination of methods is often used to collect data. The data includes **video recording** of the users including facial expressions and keystrokes and mouse movements that are logged. In addition, **a user satisfaction questionnaire** is used to find out how users actually feel about using the product. Also **structured or semi-structured interviews** may conducted with users to collect additional information.
- **Example of tasks** that are given to users include searching for information, reading different typefaces and navigating through different menus.

BINUS UNIVERSITY

People
Innovation
Excellence

USABILITY TESTING

- There are **two main performance measures** used : the time it takes typical user to complete a task and the number of errors that participants make.
- A key concern is the number of users that should be involved in a usability study : five to twelve is considered an acceptable number (Dumas and Redish,1999). But sometimes can be less if there are budget and schedule constraints.
- An example of Usability Testing : The Ipad
- Case : Budiu and Nielsen (2010) conducted a usability test of the websites and apps specific to the iPad. They wanted to understand how the interaction with the device affected people and to get feedback to their client and developers as well as people who were eager to know if the iPad lived up to the hype – which was being reported at the time it came to market.

BINUS UNIVERSITY

People
Innovation
Excellence

USABILITY TESTING

- The methods they used : usability testing with think-aloud in which users said what they were doing and thinking as they did it and an expert review.
- **A key question** they asked was : “Are user expectations different for the iPad compared with the iPhone ?”
- The usability testing carried out in two cities in United States, Fremont, California and Chicago.
- **Participants** : considered typical users and varied in age and occupation. Some of them have experienced in iPhone and had used a variety of apps
- **The test** : participant being invited to explore any application they found interesting on the iPad. They were asked to comment on what they were looking for or reading, what they liked and disliked about a side and what made it easy or difficult to carry out a task. (See Table 12.01)

USABILITY TESTING

– Table 12.01 Examples of some of the test in the iPad evaluation

Table 12.01 Some of the result test

App or website	Task
iBook	Download a free copy of *Alice's Adventures in Wonderland* and read through the first few pages.
Craigslist	Find some free mulch for your garden.
eBay	You want to buy a new iPad on eBay. Find one that you could buy from a reputable seller.
Time Magazine	Browse through the magazine and find the best pictures of the week.
Epicurious	You want to make an apple pie for tonight. Find a recipe and see what you need to buy in order to prepare it.
Kayak	You are planning a trip to Death Valley in May this year. Find a hotel located in the park or close to the park.

– The Equipment : the testing was done using the mobile usability kit shown in Fig 12.02

Figure 12.02 The mobile usability kit

People
Innovation
Excellence

USABILITY TESTING

– Usability Problem : the main findings show that the participants were able to interact with websites on the iPad but not optimal. For example : when the cover of Time Magazine appears on the iPad, it doesn't contain any hyperlinks. The content page does but it is not easily accessible. (See figure 12.03)

Figure 12.03 The cover of Time Magazine showing the contents carousel

BINUS UNIVERSITY

People
Innovation
Excellence

USABILITY TESTING

- **Interpreting and Presenting the Data** : Based on the findings of their study, Budia and Nielsen make a number of recommendation, including supporting standard navigation. **The results of the study** were written up as a report that was made publically available to app developers and the general public (www.nngroup.com).

BINUS UNIVERSITY

People
Innovation
Excellence

CONDUCTING EXPERIMENTS

- **In research contexts,** specific hypotheses are tested that make a prediction about the way users will perform with an interface.
- **The benefits** are more rigor and confidence that one interface feature is easier to understand or faster than another.
- An example of hypothesis is : **context menu** (i.e menus that provide options related to the context determined by the users previous choices) are easier to select options from compared with cascading menu.
- **Hypotheses** are often based on a theory, such as Fitts Law or previous research findings.

BINUS UNIVERSITY

People
Innovation
Excellence

CONDUCTING EXPERIMENTS

- In order to test a hypothesis ,need experimental design.
- **A concern in Experimental Design is** to determine which participants to use for which conditions in an experiment.
- There are three kind of design based on participant:
  - **Different-participant design** : a single group of participants is allocated randomly to each of the experimental conditions.
  - **Same-participant design** : all participants perform in all conditions so only half the number of participants is needed.
  - **Matched-participant design** : participants are matched in pairs based on certain user characteristics such as expertise and gender.

CONDUCTING EXPERIMENTS

- There are advantage and disadvantages of different allocations of participants to conditions (Table 12.01):

Table 12.01 Table of Comparison of different allocation participants

Design	Advantages	Disadvantages
Different Participants	No order effects	-Many participant needed. - Individual differences among participants are a problem - Can be offset to some extent by randomly assigning to groups
Same Participants	Eliminates individual differences between experimental conditions	Need to counterbalance to avoid ordering effects
Matched Participants	Same as different participants, but the effects of individual differences are reduced	Can never be sure that subjects are matched across variables that might affect performance

BINUS UNIVERSITY

People
Innovation
Excellence

FIELD STUDIES

- **Field studies** are evaluation studies that are carried out in natural settings
- **The aim** to discover how people interact with technology in the real world.
- **Field studies** that involve the deployment of prototypes or technologies in natural settings also be referred to as ‘ in the wild’.
- The field studies can range in time from just a few minutes to a period of several months or even years.
- **Data is collected primarily** by observing and interviewing people and collecting video, audio, and field notes to record what occurs in the chosen setting.
- In addition, participants may be asked to fill out paper-based or electronic diaries that run on cell phones or other handheld devices.

BINUS UNIVERSITY

People
Innovation
Excellence

INSPECTIONS

Heuristic Evaluation

- A usability inspection method that was developed by Nielsen and his colleagues.
- Guided by a set of usability principles known as heuristics, evaluate whether user-interface elements, such as dialog boxes, menus, navigation structure, online help, and so, conform to tried and tested principles.
- The original set of heuristics identified by Nielsen and his colleagues :
  - Visibility of system status
  - Match between system and the real world
  - User control and freedom
  - Consistency and standards

BINUS UNIVERSITY

People
Innovation
Excellence

INSPECTIONS

Heuristic Evaluation

- The original set of heuristics identified by Nielsen and his colleagues (continue) :
  - 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

BINUS UNIVERSITY

People
Innovation
Excellence

INSPECTIONS

Heuristic Evaluation

- Heuristics evaluation for Websites by Andy Budd(2007) :
  1. Clarity
     - Make the system as clear, concise, and meaningful as possible for the intended audience.
       - Write clear, concise copy
       - Only use technical language for a technical audience
       - Write clear and meaningful labels
       - Use meaningful icons
  2. Minimize unnecessary complexity and cognitive load
     - Make the system as simple as possible for users to accomplish their tasks
       - Remove unnecessary functionality , process step, and visual clutter
       - Use progressive disclosure to hide advanced features
       - Break down complicated processes into multiple steps
       - Prioritize using size, shape, color, alignment and proximity

BINUS UNIVERSITY

People
Innovation
Excellence

INSPECTIONS

Heuristic Evaluation

- Heuristics evaluation for Websites by Andy Budd(2007) (continue) :
  3. **Provide users with context**
     - Interfaces should provide users with a sense of context in time and space.
       - Provide a clear site name and purpose
       - Highlight the current section in the navigation
       - Provide a breadcrumb trail
       - Use appropriate feedback messages
       - Show number of steps in a process
       - Reduce perception of latency by providing visual cues (e.g. progress indicator) or by allowing users to complete other tasks while waiting.

BINUS UNIVERSITY

People
Innovation
Excellence

INSPECTIONS

Heuristic Evaluation

4. Promote a pleasurable and positive user experience
   - The user should be treated with respect and the design should be aesthetically pleasing and promote a pleasurable and rewarding experience.
     - Create a pleasurable and rewarding experience
     - Provide easily attainable goals
     - Provide rewards for usage and progression

BINUS UNIVERSITY

People
Innovation
Excellence

INSPECTIONS

Heuristic Evaluation

- Heuristic Evaluation have three stages :
  1. The briefing session , in which the experts are told what to do.
  2. The evaluation period, in which each experts typically spends 1-2 hours independently inspecting the product, using the heuristics for guidance.
  3. The debriefing session, in which the evaluators come together to discuss their findings and prioritize the problems they found and suggest solutions.

BINUS UNIVERSITY

People
Innovation
Excellence

INSPECTIONS

Cognitive Walkthroughs

- **Walkthroughs** are very focused and so are suitable for evaluating small parts of a product.
- **Cognitive walk-throughs** involve simulating how users go about problem-solving at each step in a human-computer interaction
- The steps involved in cognitive walkthroughs are :
  - The characteristics of typical users are identified and documented and sample tasks are developed that focus on the aspects of the design to be evaluated.
  - A designer and one or more expert evaluator come together to do the analysis.
  - The evaluator walk through the action sequences for each task, placing it within the context of a typical scenario.
  - As the walkthrough is being done, a record of critical information is compile .
  - The design is then revised to fix the problems presented.

BINUS UNIVERSITY

People
Innovation
Excellence

INSPECTIONS

Pluralistic Walkthroughs

- Pluralistic walkthroughs are another type of walkthrough in which users, developers, and usability experts work together to step through a task scenario, discussing usability issues associated with dialog elements involved in the scenario steps (Nielsen and Mack, 1994).
- The benefits of pluralistic walkthroughs include a strong focus on users' tasks at a detailed level , i.e. looking at the steps taken.

ANALYTICS

- **Analytics** is a method for evaluating user traffic through a system.
- When used to examine traffic on a website or part of a website , known as web analytics.
- For example : VisiStat provided Mountain Wines with data showing how their website was being used by potential customers. (See Figure 12.03 - 12.05)

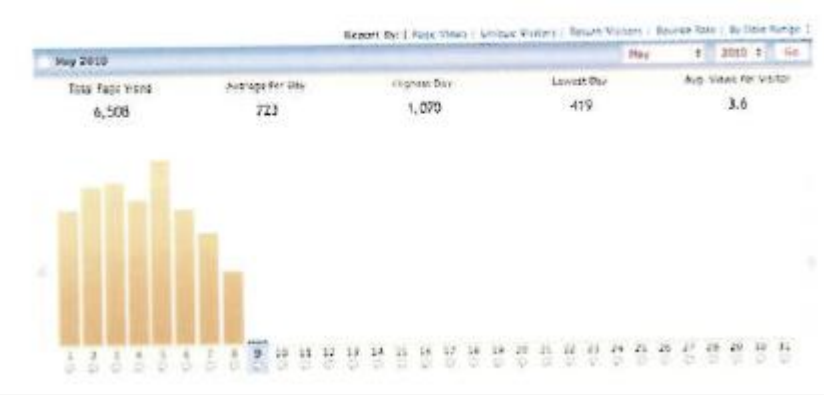

Figure 12.03 A general view of the kind of data provided by VisiStat

ANALYTICS

Figure 12.04 Clicking on May 8 provides an hourly report from midnight until 10.00 pm (only midnight and 2.00 pm – 07.00 pm shown)

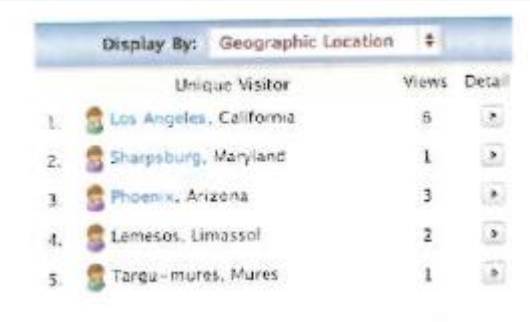

Figure 12.05 Clicking on the icon for the first hour in Figure 12.04 will shows where the IP Address of the 13 visitors to the website are located

BINUS UNIVERSITY

People
Innovation
Excellence

ANALYTICS

- **A/B Testing** is evaluate a website, part of a website, an application, or an app running on a mobile device is by carrying out a large-scale experiment to evaluate how two groups of users perform using two different designs—one of which acts as the control and the other as the experimental condition, that is, the new design being tested.

BINUS UNIVERSITY

People
Innovation
Excellence

PREDICTIVE MODELS

- Predictive models evaluate a system without users being present.
- Predictive models use formulas to derive various measures of user performance.
- Predictive modeling provides estimates of the efficiency of different systems for various kinds of task.
- The models have been used mainly are :
  - The GOMS Model

    The term GOMS is an acronym that stands for goals, operators, methods, and selection rules.

BINUS UNIVERSITY

People
Innovation
Excellence

PREDICTIVE MODELS

- The GOMS Model
  - Goals refer to a particular state the user wants to achieve (e.g. find a website on interaction design).
  - Operators refer to the cognitive process and physical actions that need to be performed in order to attain those goals.(e.g. decide on which search engine to use)

    The difference between a goal and an operator is that a goal is obtained and an operator is executed.

BINUS UNIVERSITY

People
Innovation
Excellence

PREDICTIVE MODELS

- The GOMS Model
  - Methods are learned procedures for accomplishing the goals. (e.g type in keywords in a Google search box and press the search button)
  - Selection rules are used to determine which method to select when there is more than one available for a given stage of a task. (e.g. once keywords have been entered into a search engine entry field, many search engines allow users to press the return key on the keyboard or click the go button using the mouse to progress the search.

BINUS UNIVERSITY

People
Innovation
Excellence

PREDICTIVE MODELS

- The Keystroke Level Model (KLM)
  - The KLM provide numerical predictions of user performance.
  - Tasks can be compared in terms of the time it takes to perform them when using different strategies.
  - The main benefit : that different features of systems and applications can be easily compared to see which might be the most effective for performing specific kinds of task.

PREDICTIVE MODELS

- The Keystroke Level Model (KLM)
  - When developing the KLM, Card et al (1983) analyzed the findings of many empirical studies of user performance in order to derive a standard set of approximate times for the main kinds of operators used during a task.
  - Below are the core time they proposed for these (Fig 12

Operator name	Description	Time (s)
K	Pressing a single key or button	0.35 (average)
	Skilled typist (55 wpm)	0.22
	Average typist (40 wpm)	0.28
	User unfamiliar with the keyboard	1.20
	Pressing shift or control key	0.08
P	Pointing with a mouse or other device to a target on a display	1.10
P_1	Clicking the mouse or similar device	0.20
H	Homing hands on the keyboard or other device	0.40
D	Draw a line using a mouse	Variable depending on the length of line
M	Mentally prepare to do something, e.g. make a decision	1.35
R(t)	System response time – counted only if it causes the user to wait when carrying out his/her task	*t*

Figure 12.06 The Core Time

PREDICTIVE MODELS

- The Keystroke Level Model (KLM)
  - The predicted time it takes to execute a give task is : $T_{execute} = T_K + T_P + T_H + T_D + T_M + T_R$
  - **For example** , consider how long it would take to insert the word ‘not’ into the following sentence using a word-processing program like Microsoft Word :

    “ Running after eating is normal.”

    so that becomes :

    “Running after eating is not normal”.

People
Innovation
Excellence

PREDICTIVE MODELS

- The Keystroke Level Model (KLM)
  - The times for each of these operators can then be worked out (See figure 12.07):

Mentally prepare (M)	1.35
Reach for the mouse (H)	0.40
Position mouse before the word 'normal' (P)	1.10
Click mouse (P_1)	0.20
Move hands to home position on keys (H)	0.40
Mentally prepare (M)	1.35
Type 'n' (good typist) (K)	0.22
Type 'o' (K)	0.22
Type 't' (K)	0.22
Type 'space' (K)	0.22
Total predicted time:	5.68 seconds

Figure 12.07 The List of times of these operators

  - The total predicted time is : 5.68 seconds
  - It means a duration of over 5 seconds seems a long time for inserting a word into a sentence, especially for a good typist.

PREDICTIVE MODELS

- Fitts' Law
  - Fitts' Law (Fitts, 1954) predicts the time it takes to reach a target using a pointing device.
  - The law states :

$$T = k \log_2 (D / S + 1.0)$$

Where
T = time to move the pointer to a target
D = distance between the pointer and the target
S = Size of the target
k is a constant of approximately 200 ms /bit

  - The GOMS, KLM Model and Fitts' Law can be used to predict expert, error-free performance for certain kinds of tasks.

BINUS UNIVERSITY

People
Innovation
Excellence

REFERENCES

- Interaction Design 5th Edition 2019, Chapter 15-16
- https://www.experienceux.co.uk/faqs/what-is-usability-testing/
- https://www.interaction-design.org/literature/article/heuristic-evaluation-how-to-conduct-a-heuristic-evaluation