Lecture 08: Black-Box Testing (advanced) – Security, Usability, and A/B Testing

Dietmar Pfahl
email: dietmar.pfahl@ut.ee

Spring 2019
Lectures (J. Liivi 2-111)

• Lecture 1 (14.02) – Introduction to Software Testing
• Lecture 2 (21.02) – Basic Black-Box Testing Techniques
• Lecture 3 (28.02) – BBT advanced: Combinatorial Testing
• Lecture 4 (07.03) – Basic White-Box Testing Techniques
• Lecture 5 (14.03) – Test Lifecycle, Test Tools, Test Automation
• Lecture 6 (21.03) – Test Levels / BDD & Behavior T. / GUI Testing / Visual Testing
• Lecture 7 (28.03) – BBT advanced: State-Transition Testing & Exploratory Testing
• Lecture 8 (04.04) – BBT advanced: Security, Usability and A/B Testing
• Lecture 9 (11.04) – WBT advanced: Data-Flow Testing / Mutation Testing
• Lecture 10 (18.04) – WBT advanced: Symbolic Execution, Static Code Analysis, Document Inspections, Code Review
• Lecture 11 (25.04) – Defect Estimation / Test Documentation, Organisation and Process Improvement (Test Maturity Model)
  • 02.05 - no lecture
• Lecture 12 (09.05) – Industry Guest Lecture (to be announced)
• Lecture 13 (16.05) – Exam Preparation
Structure of Lecture 8

- Lab 8
- Security Testing
- Usability Testing
- A/B Testing
Capture & Replay and Visual Testing Tools

Examples: Selenium, Sikuli

- **Good**: Simple and easy to use
- **Bad**: Test scripts must be maintained if GUI changes
‘Scriptless’ testing tool Testar

• Scripts are generated automatically (on-the-fly)
• No maintenance of scripts needed
• Main purpose: Robustness testing
  – Robustness is the ability of a computer system to cope with errors during execution and cope with erroneous input

[Link to Testar website: testar.org]
Lab 8 – Automated GUI Testing

Lab 8 (week 32: Apr 09 & 10) – Automated GUI Testing (10 points)

Lab 8 Instructions & Tools

Submission Deadlines:
- Tuesday Labs: Monday, 15 Apr, 23:59
- Wednesday Labs: Tuesday, 16 Apr, 23:59
- Penalties apply for late delivery: 50% penalty, if submitted up to 24 hours late; 100 penalty, if submitted more than 24 hours late
Tool Set-Up / Instructions

• This lab requires virtual machine
  – VirtualBox is preferred

• Import the .ova file to your VM
  – This file contains Windows 10 with Testar
  – In the desktop, you find the folder ‘Testar’ containing the Testar tool and the SUTs for the lab and homework tasks
  – Make sure you have enough free space (>= 20 GB)

Please set up before the lab starts!
Testar tool (1)

- GUI contains graphical objects called widgets
- Widgets form a hierarchy called widget tree
- Widgets have properties (title, etc.)
- Widget tree and properties of each widget form a GUI state
- GUI state changes, if user executes an action (click, write, etc.)
Testar tool (2)

• Example widget of Calculator program
Testar tool (3)

- Widget tree of Calculator program
Testar tool (4) – What it can do ...

Invalid input to SUT:
1) error message (specified) → NOT A FAILURE → can be ignored (if necessary apply Filter to avoid it’s activation)
2) error message (not specified) → FAILURE → define Oracle to detect it
3) no error message → FAILURE (BUT: Testar cannot detect it)

Valid input to SUT:
1) no error message → NOT A FAILURE, i.e., common behavior of the SUT and we don't care about it → apply filters to avoid unwanted behavior (e.g., ‘close’ button)
2) error message (not specified) → FAILURE --> define Oracle to detect it
3) error message (specified) → FAILURE (BUT: Testar cannot detect it because no Oracle should be defined for specified messages)
Testar tool (5) – What it can do ...

Testar always detects:
• ... when SUT crashes or hangs (no Oracle needs to be defined)

Testar cannot detect:
• ... wrong business logic implemented (e.g., $2 + 2 \Rightarrow \text{result} == 5$)
• ... when specified error messages are shown although the input to a function was valid
• ... when invalid input is not handled properly (i.e., no specified error message is presented) AND there is no (unspecified) exception thrown or a program crash/hanging
Testar tool (6)

1. Start SUT
2. Scan GUI state
3. Check oracles*
   - Failure detected
   - No failure detected
4. Check stop criteria
   - Yes
     - Stop SUT
   - No
5. Derive (+filter) actions
6. Select action
7. Execute action
8. * or crash
Testar tool (6)

Some parameters have to be set before running Testar (-> stop criteria)

Start SUT

Scan GUI state

Check oracles*

Check stop criteria

Derive (+filter) actions

Select action

Execute action

No

Failure detected

Stop SUT

Yes

No failure detected

Sequence leading to failure

* or crash

Use Regular Expressions to define
• Oracles
• Filters

Some parameters have to be set before running Testar (-> stop criteria)
Testar tool (7) – Stop criteria

- Number of Sequences (=Iterations)
- Number of Actions per Sequence
  - Identify when there have enough actions been performed for one test

1 2 1 + 4 4 4 * 3 = Sequence 1: 10 actions
+ – 3 ^ 4 / 0 4 = 2 Sequence 2: 10 actions
...
Testar tool (8) – Regular expressions

• Example:

```
\b[A-Z0-9._%+-]+@[A-Z0-9.-]+\.[A-Z]{2,}\b
```
Testar tool (9) – Oracles

• Testar can detect two types of failures:
  – When SUT crashes or hangs (no oracles needed)
  – When SUT’s widget title matches a specified oracle (regex)

Example: Test engineer specifies „fail“ as an oracle
• If the word „fail“ appears in some widget title, then Testar reports it as a failure (test = fail)
• If the word “fail” doesn’t appear in a sequence, then by default the test passes (test = pass)
Testar tool (10) - Filters

• Default: Testar executes actions on all the widgets it finds
  – Including buttons on the title bar (e.g., close, minimize, maximize, system)

• User can define that specific widgets shall not be executed
  – For example, forbid clicking on ‘close’ or ‘minimize’ buttons
  – To do this, an action filter can be configured
    • Filtering is done by matching the widget title to defined filters (using regex)
    • For example, user can specify „.*close.*“ as a filter and Testar will not execute actions on widgets, where title contains „close“
Testar tool (11) – Filter example

Example:

`.*Backspace.*l.*CE.*l.*View.*`

This expression could be used to filter out (=ignore) all clicks to control elements whose title contain one of the given strings.
Testar output ➞ folder structure

- **Testar**
  - **Bin**
    - **Testar.bat** ➞ File to run Testar
  - **Output**
    - **Sequences_suspiciousTitle** ➞ Sequences that matched oracles
    - **Sequences_unexpectedClose** ➞ Sequences where the SUT closed unexpectedly
  - **Graphs** ➞ Contains raw graph files and a converter

- **SUT**
  - **Calculator** ➞ Exercise in lab
  - **ATM** ➞ Homework task

Note: This slide only shows the most important folders.
Interpreting output (1)

- Tester must determine if the output reports a …
  - true positive -> an actual failure or ...
  - false positive -> reported as failure, although it is not
  - That’s why the folders are called „unexpected…“ and „suspicious…“

- Tester must look through all the sequences
- Beware that there are many failure duplicates, and there is no (easy) way to automatically analyse the output files
Interpreting output (2)

- Testar generates files that can be converted into graphs
  - "offline_graph_conversion.bat" is automatically provided in each graph folder
- All the graphs are converted automatically using the .bat file
  - you use the screenshotted graph to understand what happened in the sequence
- Looking through the screenshotted graphs is simple but …
  - it can happen that there are some layout issues and not all GUI states in a screenshot are connected
• Like in this example, where you can see only one state before failure

Which was previous state?
• To find the previous state, use action number in brackets
• Search for lower number to find previous state
• Ctrl + f
• It will show you the previous action
• This way you can move through the graph if some parts are not connected
Hints (1)

• Use a reasonable number of sequences (e.g., something between 20 and 100)
  • If the number is too small, you might not trigger a (potential) failure at all
  • If the number is too high, the execution might take too long and the (manual) analysis of the output files might take too much time

• Use a reasonable number of actions in each sequence
  • It will keep graphs smaller and therefore easier to read
  • A maximum of 20 actions should be fine but what to choose depends also on how you use the action filters
Hints (2)

• If you keep finding the same failure many times, you should consider filtering out the path leading to that failure
  • It will save you time while inspecting graphs
  • Also, make sure, that you don’t apply filters too early, otherwise you might miss some of the failures

• Information messages are not to be considered failures if they match those listed in the appendixes A (in-lab task) and C (homework task)
  • Don’t write oracles matching these messages

• Check out Testar manual for more information about the tool
Structure of Lecture 8

- Lab 8
- Security Testing
- Usability Testing
- A/B Testing
Software Product Quality Model
– ISO 25010 Standard

Security
Security Testing – What?

• Security Testing = Type of testing that ...
  • intends to **uncover vulnerabilities** of the system and
  • determine that its data and resources are **protected from possible intruders**

• Focus Areas:
  • **Network security**: This involves looking for vulnerabilities in the network infrastructure (resources and policies).
  • **System software security**: This involves assessing weaknesses in the various software (operating system, database system, and other software) the application depends on.
  • **Client-side application security**: This deals with ensuring that the client (browser or any such tool) cannot be manipulated.
  • **Server-side application security**: This involves making sure that the server code and its technologies are robust enough to fend off any intrusion.
Security Testing – Why?

• Primary purpose of security testing is to identify the vulnerabilities and subsequently repairing them

• Security Testing helps in
  • improving the current system
  • ensuring that the system will work for longer time
  • finding out loopholes that can cause loss of important information
Security Testing – Basic Concepts

• Confidentiality
• Integrity
• Authentication
• Authorization
• Accountability
• Non-repudiation
• [Availability]
Security Testing – Basic Concepts

- **Confidentiality**
- Integrity
- Authentication
- Authorization
- Accountability
- Non-repudiation
- [Availability]

is about …

... ensuring that information is accessible only for those with authorized access and … preventing information theft
Security Testing – Basic Concepts

- Confidentiality
- **Integrity**
- Authentication
- Authorization
- Accountability
- Non-repudiation
- [Availability]

is about …

... allowing the receiver of data to determine that the information which it is providing is correct
Security Testing – Basic Concepts

- Confidentiality
- Integrity
- **Authentication**
- Authorization
- Accountability
- Non-repudiation
- [Availability]

is about ...
... establishing the identity of the user
Security Testing – Basic Concepts

• Confidentiality
• Integrity
• Authentication
• **Authorization**
• Accountability
• Non-repudiation
• [Availability]

is about …
… determining that a data requester is allowed to receive a service or perform an operation
Security Testing – Basic Concepts

- Confidentiality
- Integrity
- Authentication
- Authorization
- **Accountability**
- Non-repudiation
- [Availability]

... is about ...

... taking responsibility (and transparency about the responsibilities taken) for establishing security
Security Testing – Basic Concepts

• Confidentiality
• Integrity
• Authentication
• Authorization
• Accountability
• **Non-repudiation**
• [Availability]

is about …

... preventing the later denial that an action happened or a communication took place, etc.
Security Testing – Basic Concepts

• Confidentiality
• Integrity
• Authentication
• Authorization
• Accountability
• Non-repudiation
• [Availability]

is about …
… assuring that information and communications services will be ready for use when expected/needed
Security Testing – What?

Main security testing practices include:
• Vulnerability Scanning
• Security Scanning
• Penetration Testing
• Ethical Hacking
• Risk Assessment
• Security Auditing
• Posture Assessment
• Password cracking
Security Testing – What?

Main security testing practices include:

- **Vulnerability Scanning**
  - Involves scanning of the application for all known vulnerabilities
  - Typically done with the help of vulnerability scanning software, e.g., Nessus, Nikto, Gendarme, Flawfinder, etc.

- Security Scanning
- Penetration Testing
- Ethical Hacking
- Risk Assessment
- Security Auditing
- Posture Assessment
- Password cracking
Security Testing – What?

Main security testing practices include:

• **Vulnerability Scanning**
• **Security Scanning**
• Penetration Testing
• Ethical Hacking
• Risk Assessment
• Security Auditing
• Posture Assessment
• Password cracking

Involves scanning the usage of a system in operation to monitor whether attacks on the system are being tried or perhaps prepared.
Security Testing – What?

Main security testing practices include:

- Vulnerability Scanning
- Security Scanning
- **Penetration Testing**
- Ethical Hacking
- Risk Assessment
- Security Auditing
- Posture Assessment
- Password cracking

Testers try to (unauthorized) enter into the application with the help of another software system and via (combinations of) loopholes that the application has kept open unknowingly.
Security Testing – What?

Main security testing practices include:

- Vulnerability Scanning
- Security Scanning
- Penetration Testing
- **Ethical Hacking**
- Risk Assessment
- Security Auditing
- Posture Assessment
- Password cracking

Involves the performance of penetration tests over the wide network on the system under test

Conducted by ethical hackers who try to find possible problems in the system
Security Testing – What?

Main security testing practices include:
• Vulnerability Scanning
• Security Scanning
• Penetration Testing
• Ethical Hacking
• **Risk Assessment**
• Security Auditing
• Posture Assessment
• Password cracking

Analyzing and deriving the risk as a function of potential damage (or loss) and the possibility of damage occurrence
Typically carried out in the form of scenario analysis, interviews, and expert discussions
Security Testing – What?

Main security testing practices include:

- Vulnerability Scanning
- Security Scanning
- Penetration Testing
- Ethical Hacking
- Risk Assessment
- **Security Auditing**
- Posture Assessment
- Password cracking

**Security Auditing** involves hands on internal inspection of Operating Systems and Applications, often via line-by-line inspection of the code.

A security audit is a systematic evaluation of the security of a software system used by a company.
Security Testing – What?

Main security testing practices include:
• Vulnerability Scanning
• Security Scanning
• Penetration Testing
• Ethical Hacking
• Risk Assessment
• Security Auditing
• **Posture Assessment**
• Password Cracking

combines Security Scanning, Ethical Hacking and Risk Assessments to show an overall Security Posture (=attitude towards security) of the organization using a software system
Security Testing – What?

Main security testing practices include:

- Vulnerability Scanning
- Security Scanning
- Penetration Testing
- Ethical Hacking
- Risk Assessment
- Security Auditing
- Posture Assessment
- **Password Cracking**

Password cracking programs can be used to identify weak passwords. Password cracking verifies that users are employing sufficiently strong passwords.
Security Testing – What?

Example of a basic security test:

- Log into the web application.
- Log out of the web application.
- Click the BACK button of the browser (Check if you are asked to log in again or if you are provided the logged-in application.)

- Most types of security testing involve complex steps and out-of-the-box thinking but, sometimes, it is simple tests like the one above that help expose the most severe security risks.


OWASP Top 10 security threats for 2013 were:
- Injection
- Broken Authentication and Session Management
- Cross-Site Scripting (XSS)
- Insecure Direct Object References
- Security Misconfiguration
- Sensitive Data Exposure
- Missing Function Level Access Control
- Cross-Site Request Forgery (CSRF)
- Using Known Vulnerable Components
- Unvalidated Redirects and Forwards
The Open Web Application Security Project (OWASP) is a great resource for software security professionals. Be sure to check out the Testing Guide: https://www.owasp.org/index.php/Category:OWASP_Testing_Project

OWASP Top 10 security threats for 2013 were:
- Injection
- Broken Authentication and Session Management
- Cross-Site Scripting (XSS)
- Insecure Direct Object References
- Security Misconfiguration
- Sensitive Data Exposure
- Missing Function Level Access Control
- Cross-Site Request Forgery (CSRF)
- Using Known Vulnerable Components
- Unvalidated Redirects and Forwards
How to avoid SQL injection vulnerability?

Instead of:

```java
String query = "SELECT * FROM Users WHERE Username= "
    + request.getParameter("username")
    + "AND Password= "
    + request.getParameter("password");

try {
    Statement statement = connection.createStatement();
    ResultSet results = statement.executeQuery(query);
}
```

Which might result in a SQL query string like this:

```
SELECT * FROM Users WHERE Username='1' OR '1' = '1' AND Password='1' OR '1' = '1'
```
How to avoid SQL injection vulnerability?

Use java ‘prepared statement’:

```java
String username = request.getParameter("username");
String password = request.getParameter("password");
// perform input validation to detect attacks

String query = "SELECT * FROM Users WHERE Username= ? AND Password= ?";
PreparedStatement pstmt = connection.prepareStatement(query);
pstmt.setString( 1, username);
pstmt.setString( 2, password);
ResultSet results = pstmt.executeQuery();

Example with Hibernate Query Language (HQL) can be found here: https://www.owasp.org/index.php/SQL_Injection_Prevention_Cheat_Sheet
```
Structure of Lecture 8

• Lab 8
• Security Testing
• Usability Testing
• A/B Testing
Software Product Quality Model – ISO 25010 Standard

Usability
# Testing Usability Requirements

<table>
<thead>
<tr>
<th>Problem counts</th>
<th>Task time</th>
<th>Keystroke counts</th>
<th>Opinion poll</th>
<th>Score for understanding</th>
<th>Design-level reqs</th>
<th>Product-level reqs</th>
<th>Guideline adherence</th>
<th>Development process reqs</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1: At most 1 of 5 novices shall encounter critical problems during tasks Q and R. At most 5 medium problems on list.</td>
<td>R2: Novice users shall perform tasks Q and R in 15 minutes. Experienced users tasks Q, R, S in 2 minutes.</td>
<td>R3: Recording breakfast shall be possible. No mouse.</td>
<td>R4: 80% of users shall find system easy to recommend to others.</td>
<td>R5: Show 5 users 10 common error messages. Ask for the cause. 80% of these should be very large.</td>
<td>R6: System shall use screen pictures in app. xx, buttons work as app. yy.</td>
<td>R7: For all code fields, user shall be able to select value from drop-down list.</td>
<td>R8: System shall follow style guide zz. Menus shall have at most three levels.</td>
<td>R9: Three prototype versions shall be made and usability tested during design.</td>
</tr>
</tbody>
</table>
Testing Usability Requirements

How to test:
- Define several (typical) usage scenarios involving tasks Q and R
- Select test users and classify as 'novice' and 'experienced'
- Let 5 (or better 10, 15) novices perform the scenarios
- Observe what problems they encounter
- Classify and count observed problems

<table>
<thead>
<tr>
<th>Problem counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1: At most 1 of 5 novices shall encounter critical problems during tasks Q and R. At most 5 medium problems on list.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task time</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2: Novice users shall perform tasks Q and R in 15 minutes. Experienced users tasks Q, R, S in 2 minutes.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Keystroke counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>R3: Recording breakfast shall be possible with 5 keystrokes per guest. No mouse.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opinion poll</th>
</tr>
</thead>
<tbody>
<tr>
<td>R4: 80% of users shall find system easy to learn. 60% shall recommend system to others.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score for understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>R5: Show 5 users 10 common error messages, e.g. Amount too large. Ask for the cause. 80% of the answers shall be correct.</td>
</tr>
</tbody>
</table>
Usability Test Types + Environment

Rubin’s Types of Usability Tests (Rubin, 1994, p. 31-46)

Exploratory test – early product development

Assessment test – most typical, either early or midway in the product development

Validation test – confirmation of product’s usability

Comparison test – compare two or more designs; can be used with other three types of tests

Think aloud

Observe

Audio/video recording

Rubin’s Types of Usability Tests (Rubin, 1994, p. 31-46)

Exploratory test – early product development

Assessment test – most typical, either early or midway in the product development

Validation test – confirmation of product’s usability

Comparison test – compare two or more designs; can be used with other three types of tests
Usability Testing – What? How?

- **Test Focus**
  - **Understandability**
    - Easy to understand?
  - **Ease of learning**
    - Easy to learn?
  - **Operability**
    - Matches purpose & environment of operation?
    - Ergonomics: color, font, sound,...
  - **Communicativeness**
    - In accordance with psychological characteristics of user?

- **Test Environments**
  - Free form tasks
  - Procedure scripts
  - Paper screens
  - Mock-ups
  - Field trial
Evaluating UI Designs

- Inspection Methods
  - Heuristic Evaluation
  - Cognitive Walkthrough
- Usability Testing
  - Laboratory Experiment
  - Guidelines Review
  - Field Study
Evaluating UI Designs

Inspection Methods
- Heuristic Evaluation
  - Cognitive Walkthrough
  - Guidelines Review

Usability Testing
- Laboratory Experiment
- Field Study

Cheap
Expensive
Evaluating UI Designs

- **Inspection Methods**
  - **Heuristic Evaluation**
    - 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
    - Scenarios may be used to guide analysis

- **Usability Testing**
  - **Field Study**
    - Laboratory Experiment
    - Review

---

**Cognitive Walkthrough**
Evaluating UI Designs

Inspection Methods

Heuristic Evaluation

Cognitive Walkthrough

Guidelines

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

Usability Testing

Laboratory Experiment

Field Study
Heuristic Evaluation by Inspection

List of 10 Heuristics according to (Nielsen, 2005):

<table>
<thead>
<tr>
<th></th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visibility of system status</td>
</tr>
<tr>
<td>2</td>
<td>Match between the system and the real world</td>
</tr>
<tr>
<td>3</td>
<td>User control and freedom</td>
</tr>
<tr>
<td>4</td>
<td>Consistency and standards</td>
</tr>
<tr>
<td>5</td>
<td>Error prevention</td>
</tr>
<tr>
<td>6</td>
<td>Recognition rather than recall</td>
</tr>
<tr>
<td>7</td>
<td>Flexibility and efficiency of use</td>
</tr>
<tr>
<td>8</td>
<td>Aesthetic and minimalist design</td>
</tr>
<tr>
<td>9</td>
<td>Help users recognize, diagnose, and recover from errors</td>
</tr>
<tr>
<td>10</td>
<td>Help and documentation</td>
</tr>
</tbody>
</table>

List violations of heuristics:

Rank by severity: 0...4
0: positive (or neutral) aspect of system
...4: major, catastrophic aspect of system


### Heuristic Evaluation by Inspection

#### List of 10 Heuristics according to (Nielsen, 2005):

<table>
<thead>
<tr>
<th></th>
<th>Visibility of system status</th>
<th>Match between the system and the real world</th>
<th>User control and freedom</th>
<th>Consistency and standards</th>
<th>Error prevention</th>
<th>Recognition rather than recall</th>
<th>Flexibility and efficiency of use</th>
<th>Aesthetic and minimalist design</th>
<th>Help users recognize, diagnose, and recover from errors</th>
<th>Help and documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visibility of system status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Match between the system and the real world</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>User control and freedom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Consistency and standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Error prevention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Recognition rather than recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Flexibility and efficiency of use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Aesthetic and minimalist design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Help users recognize, diagnose, and recover from errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Help and documentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Visibility of system status:

At all times, the system should inform the end user what is currently ongoing, e.g., via timely and adequate feedback.

#### List violations of heuristics:

**Rank by severity: 0...4**

- **0**: positive (or neutral) aspect of system
- **...**
- **4**: major, catastrophic aspect of system

---


Heuristic Evaluation

List of 10 Heuristics according to (Nielsen, 2005):

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visibility of system status</td>
</tr>
<tr>
<td>2</td>
<td>Match between the system and the real world</td>
</tr>
<tr>
<td>3</td>
<td>User control and freedom</td>
</tr>
<tr>
<td>4</td>
<td>Consistency and standards</td>
</tr>
<tr>
<td>5</td>
<td>Error prevention</td>
</tr>
<tr>
<td>6</td>
<td>Recognition rather than recall</td>
</tr>
<tr>
<td>7</td>
<td>Flexibility and efficiency of use</td>
</tr>
<tr>
<td>8</td>
<td>Aesthetic and minimalist design</td>
</tr>
<tr>
<td>9</td>
<td>Help users recognize, diagnose, and recover from errors</td>
</tr>
<tr>
<td>10</td>
<td>Help and documentation</td>
</tr>
</tbody>
</table>

Match between the system and the real world:

The system should speak the language of the end user (i.e., it should use adequate terminology from the domain of the end user).

The system should follow the conventions of the real world and present information in a logic and well-structured form.


### Heuristic Evaluation

#### List of 10 Heuristics according to (Nielsen, 2005):

<table>
<thead>
<tr>
<th>Rank</th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visibility of system status</td>
</tr>
<tr>
<td>2</td>
<td>Match between the system and the real world</td>
</tr>
<tr>
<td>3</td>
<td>User control and freedom</td>
</tr>
<tr>
<td>4</td>
<td>Consistency and standards</td>
</tr>
<tr>
<td>5</td>
<td>Error prevention</td>
</tr>
<tr>
<td>6</td>
<td>Recognition rather than recall</td>
</tr>
<tr>
<td>7</td>
<td>Flexibility and efficiency of use</td>
</tr>
<tr>
<td>8</td>
<td>Aesthetic and minimalist design</td>
</tr>
<tr>
<td>9</td>
<td>Help users recognize, diagnose, and recover from errors</td>
</tr>
<tr>
<td>10</td>
<td>Help and documentation</td>
</tr>
</tbody>
</table>

#### User control and freedom:

End users often select a certain function by accident; in that case, they should be able to easily find a way out of this unwanted system state, e.g., by simply choosing ‘undo’ or ‘redo’ functionality.

#### Rank by severity: 0...4

- **0**: positive (or neutral) aspect of system
- **...**
- **4**: major, catastrophic aspect of system

---


### Heuristic Evaluation by Inspection

#### List of 10 Heuristics according to (Nielsen, 2005):

<table>
<thead>
<tr>
<th>Rank</th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visibility of system status</td>
</tr>
<tr>
<td>2</td>
<td>Match between the system and the real world</td>
</tr>
<tr>
<td>3</td>
<td>User control and freedom</td>
</tr>
<tr>
<td>4</td>
<td>Consistency and standards</td>
</tr>
<tr>
<td>5</td>
<td>Error prevention</td>
</tr>
<tr>
<td>6</td>
<td>Recognition rather than recall</td>
</tr>
<tr>
<td>7</td>
<td>Flexibility and efficiency of use</td>
</tr>
<tr>
<td>8</td>
<td>Aesthetic and minimalist design</td>
</tr>
<tr>
<td>9</td>
<td>Help users recognize, diagnose, and recover from errors</td>
</tr>
<tr>
<td>10</td>
<td>Help and documentation</td>
</tr>
</tbody>
</table>

**Consistency and standards:**

End users should not have to ask themselves whether different words (or symbols) and actions mean the same or do the same. Conventions that are typical for the used platform should be honored.

0: positive (or neutral) aspect of system

...  

4: major, catastrophic aspect of system

---


**Heuristic Evaluation**

List of 10 Heuristics according to (Nielsen, 2005):

<table>
<thead>
<tr>
<th></th>
<th>Visibility of system status</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Match between the system and the real world</td>
</tr>
<tr>
<td>3</td>
<td>User control and freedom</td>
</tr>
<tr>
<td>4</td>
<td>Consistency and standards</td>
</tr>
<tr>
<td>5</td>
<td>Error prevention</td>
</tr>
<tr>
<td>6</td>
<td>Recognition rather than recall</td>
</tr>
<tr>
<td>7</td>
<td>Flexibility and efficiency of use</td>
</tr>
<tr>
<td>8</td>
<td>Aesthetic and minimalist design</td>
</tr>
<tr>
<td>9</td>
<td>Help users recognize, diagnose, and recover from errors</td>
</tr>
<tr>
<td>10</td>
<td>Help and documentation</td>
</tr>
</tbody>
</table>

**Error prevention:**

Even better than error messages is it to design the UI such that no errors occur, i.e., situations in which the end user might make a mistake should be avoided.

Before a transaction with potentially incorrect end user input is executed, a dialog seeking confirmation is useful.


# Heuristic Evaluation by Inspection

**List of 10 Heuristics according to (Nielsen, 2005):**

<table>
<thead>
<tr>
<th>No.</th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visibility of system status</td>
</tr>
<tr>
<td>2</td>
<td>Match between the system and the real world</td>
</tr>
<tr>
<td>3</td>
<td>User control and freedom</td>
</tr>
<tr>
<td>4</td>
<td>Consistency and standards</td>
</tr>
<tr>
<td>5</td>
<td>Error prevention</td>
</tr>
<tr>
<td>6</td>
<td>Recognition rather than recall</td>
</tr>
<tr>
<td>7</td>
<td>Flexibility and efficiency of use</td>
</tr>
<tr>
<td>8</td>
<td>Aesthetic and minimalist design</td>
</tr>
<tr>
<td>9</td>
<td>Help users recognize, diagnose, and recover from errors</td>
</tr>
<tr>
<td>10</td>
<td>Help and documentation</td>
</tr>
</tbody>
</table>

**Recognition rather than recall:**

The cognitive load of the end user can be reduced by making objects, actions, and options visible and recognizable.

End users should not be forced to remember information when moving from one dialog step to the next.

Help text should always be easily visible or, at least, easy to find.
### Heuristic Evaluation by Inspection

#### List of 10 Heuristics according to (Nielsen, 2005):

<table>
<thead>
<tr>
<th></th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visibility of system status</td>
</tr>
<tr>
<td>2</td>
<td>Match between the system and the real world</td>
</tr>
<tr>
<td>3</td>
<td>User control and freedom</td>
</tr>
<tr>
<td>4</td>
<td>Consistency and standards</td>
</tr>
<tr>
<td>5</td>
<td>Error prevention</td>
</tr>
<tr>
<td>6</td>
<td>Recognition rather than recall</td>
</tr>
<tr>
<td>7</td>
<td>Flexibility and efficiency of use</td>
</tr>
<tr>
<td>8</td>
<td>Aesthetic and minimalist design</td>
</tr>
<tr>
<td>9</td>
<td>Help users recognize, diagnose, and recover from errors</td>
</tr>
<tr>
<td>10</td>
<td>Help and documentation</td>
</tr>
</tbody>
</table>

#### Flexibility and efficiency of use:

- Shortcuts and special keys – often not known by novice users – can be helpful to expert users and make their interaction with the system more efficient.

The system should be usable according to the needs of different user groups (flexibility).
### Heuristic Evaluation

**List of 10 Heuristics according to (Nielsen, 2005):**

<table>
<thead>
<tr>
<th></th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visibility of system status</td>
</tr>
<tr>
<td>2</td>
<td>Match between the system and the real world</td>
</tr>
<tr>
<td>3</td>
<td>User control and freedom</td>
</tr>
<tr>
<td>4</td>
<td>Consistency and standards</td>
</tr>
<tr>
<td>5</td>
<td>Error prevention</td>
</tr>
<tr>
<td>6</td>
<td>Recognition rather than recall</td>
</tr>
<tr>
<td>7</td>
<td>Flexibility and efficiency of use</td>
</tr>
<tr>
<td>8</td>
<td>Aesthetic and minimalist design</td>
</tr>
<tr>
<td>9</td>
<td>Help users recognize, diagnose, and recover from errors</td>
</tr>
<tr>
<td>10</td>
<td>Help and documentation</td>
</tr>
</tbody>
</table>

### Aesthetic and minimalist design:

(Standard) dialogs shouldn’t contain any irrelevant or rarely needed information. Every unnecessary information competes for the attention of the end user with the relevant information and thus diminishes its visibility.

**Aesthetics:** color, alignment, contrast, proximity

---

# Heuristic Evaluation

List of 10 Heuristics according to Nielsen (2005):

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Rank by severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Visibility of system status</td>
<td>0...4</td>
</tr>
<tr>
<td>2 Match between the system and the real world</td>
<td></td>
</tr>
<tr>
<td>3 User control and freedom</td>
<td></td>
</tr>
<tr>
<td>4 Consistency and standards</td>
<td></td>
</tr>
<tr>
<td>5 Error prevention</td>
<td></td>
</tr>
<tr>
<td>6 Recognition rather than recall</td>
<td></td>
</tr>
<tr>
<td>7 Flexibility and efficiency of use</td>
<td></td>
</tr>
<tr>
<td>8 Aesthetic and minimalist design</td>
<td></td>
</tr>
<tr>
<td>9 Help users recognize, diagnose, and recover from errors</td>
<td></td>
</tr>
<tr>
<td>10 Help and documentation</td>
<td></td>
</tr>
</tbody>
</table>

Help users recognize, diagnose, and recover from errors:

Error messages should use simple language, not contain cryptic codes, clearly identify the problem, and make a constructive solution to solve the problem.


Help and documentation:

Ideally, a system should be usable without documentation (self-explanatory), sometimes it is useful to offer help.

Help texts and documentation should be easy to find and search.

Help texts and documentation should focus on the end users’ (typical) needs and offer concrete steps on how to use the system for a specific task.
Evaluating UI Designs

- Inspection Methods
  - Heuristic Evaluation
    - Cognitive Walkthrough
  - Guidelines Review
- Usability Testing
  - Laboratory Experiment
  - Field Study
Evaluating UI Designs

Written guidelines recommended for larger projects:
- Screen layout
- Appearance of objects
- Terminology
- Wording of prompts and error messages
- Menu’s
- Direct manipulation actions and feedback
- On-line help and other documentation
Guidelines for Screen Layout & Appearance

P.A.R.C.

- Proximity
- Alignment
- Repetition
- Contrast
Guidelines for Screen Layout & Appearance

P.A.R.C.

Proximity
Objects that are close to another (related) appear to form groups.

Alignment

Repetition

Contrast
Guidelines for Screen Layout & Appearance

P.A.R.C.

- Proximity
- Alignment
- Repetition
- Contrast

Alignment of objects generates clarity / avoids confusion.

Alignment in alignment:

\[ \overbrace{\cdot-o-o-o-o-o-o-o-o-o-o-o-o-o-o} \]

Alignment out of alignment:

\[ \overbrace{\cdot-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o} \]
Guidelines for Screen Layout & Appearance

P.A.R.C.

Proximity

Alignment

Repetition

Contrast

Visual aspects such as color, shape and size link (similar) items together.
Guidelines for Screen Layout & Appearance

P.A.R.C.

- Proximity
- Alignment
- Repetition
- Contrast

Contrast makes differences easier recognizable.
Guidelines for Screen Layout & Appearance

P.A.R.C.

Proximity

Alignment

Repetition

Contrast

Contrast makes differences easier recognizable. Use: size, color, value, shape, position, direction, …

Contrast makes differences easier recognizable. Use: size, color, value, shape, position, direction, …
Evaluating UI Designs

Inspection Methods

- Heuristic Evaluation
- Cognitive Walkthrough
- Guidelines Review

Usability Testing

- Laboratory Experiment
- Field Study
Evaluating UI Designs

Inspection Methods

Heuristic Evaluation

Usability Testing

Laboratory Experiment

Field Study

Usability testing in a controlled environment
- There is a test set of users
- They perform pre-specified tasks
- Data is collected (quantitative and qualitative)
- Take mean and/or median value of measured attributes
- Compare to goal or another system
Evaluating UI Designs

1. Direct observation in actual use
   - discover new uses
   - take notes, don’t help, chat later

2. Logging actual use
   - objective, not intrusive
   - great for identifying errors
   - which features are/are not used
   - privacy concerns

- Cognitive Walkthrough
- Guidelines Review
- Usability Testing
- Laboratory Experiment
- Field Study
Evaluating UI Designs

3. Questionnaires and interviews with real users
   ask users to recall critical incidents
   questionnaires must be short and easy to return
4. Focus groups
   6-9 users
   skilled moderator with pre-planned script
   computer conferencing??
5. On-line direct feedback mechanisms
   initiated by users
   may signal change in user needs
   trust but verify
6. Bulletin boards and user groups
UI Design & Usability

Recommendation for those who want to know more about UI Design (and Usability):

• Course: LTAT.05.007 - Human Computer Interaction
• Teacher: Alexander Nolte
Structure of Lecture 8

• Lab 8
• Security Testing
• Usability Testing
• A/B Testing
A/B Testing

Two GUI Versions A & B
A/B Testing (cont’d)

Visual Website Optimizer divides traffic between the two versions

Tool support

Randomly selected

Two GUI Versions A & B
A/B Testing (cont’d)

Tools:
https://blog.crazyegg.com/2014/06/25/best-testing-software/

More Sales?

Two GUI Versions A & B
A/B Testing – Real-World Example

Former US president Obama’s 2008 Election campaign

A/B Testing – Real-World Example

Button variations:
A/B Testing – Real-World Example

Media variations:

- Family Image
- Change Image
- Barack’s Video
- …
(6 alternatives in total)
A/B Testing – Real-World Example

In total

4 x 6 = 24 combinations

(including the original button and medium)

Sign-up rates for each section (button, medium)
A/B Testing – Real-World Example

In total

4 x 6 = 24 combinations

(including the original button and medium)

Combination 11: “Learn More” & “Family Image”

Sign-up rates for each <button, medium>-combination
A/B Testing – Real-World Example

Former US president Obama’s 2008 Election campaign

Combination 11: “Learn More” & “Family Image”

A/B Testing – Multivariate Testing

- Only 2 items in previous example
A/B Testing (cont’d)

What to vary ...

• Call-To-Actions – Placement, wording, size
• Copywriting – Value propositions, product descriptions
• Forms – Their length, field types, text on the forms.
• Layout – Homepage, content pages, landing pages
• Product pricing – Try testing for revenue by testing your prices
• Images/Videos – Their placement, content and size
• Amount of content on the page (short vs. long)

Link:

http://conversionxl.com/how-to-build-a-strong-ab-testing-plan-that-gets-results/
A/B Testing -- Tools

A/B Split Test Significance Calculator by VWO
• A widely used tool for calculating the significance of your A/B testing results.

A/B Split and Multivariate Test Duration Calculator by VWO
• The calculator allows you to calculate maximum duration for which your test should run.

Crazyegg, Inspectlet, Clicktale and Mouseflow
• Heatmap software for tracking your visitor’s behavior on your site. You can get good data for hypotheses generation.
Next Week

• Quiz 8

• Lab 8:
  – Automated GUI Testing

• Lecture 9:
  White-Box Testing (advanced):
  – Data-Flow Testing & Mutation Testing