An Empirical Study on the Efficiency of Graphical vs. Textual Representations in Requirements Comprehension

Zohreh Sharafi, Alessandro Marchetto, Angelo Susi, Giuliano Antoniol and Yann-Gaël Guéhéneuc

ICPC’13
Outline

- Introduction
- Problem and Motivation
- Related Work
- Our Empirical Study
  - Design
  - Result and Analysis
  - Threats to Validity
- Conclusion
Introduction

Documentation is often only textual or graphical

Comments
Manual pages

Requirements

Graphical Documentation

Not always available

Program comprehension relies heavily on documentation
Introduction

Is there an impact (accuracy, time) of the kind of document representation on program comprehension?

Requirements

☐ Structured text only
☐ Graphical representation only
☐ Mixed textual and graphical representation

Is there an impact of the mother language, degree of study, or gender?
# Requirement Representation

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Textual</th>
<th>Graphical (TROPOS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td>Higher accuracy or no difference?</td>
</tr>
<tr>
<td>![Checkmark]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td></td>
<td>More time spent on textual or more on graphical?</td>
</tr>
<tr>
<td>![Hourglass]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Effort</strong></td>
<td></td>
<td>Visual effort?</td>
</tr>
<tr>
<td>![Golden Sphere]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Gender Preference

- Does gender affects the preferred document representation?

- Does gender and representation affect program comprehension?
## Related Works

<table>
<thead>
<tr>
<th>Authors</th>
<th>Representation</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ottensooser et al. [3]</td>
<td>Textual representations</td>
<td>Significant improvement in understanding of business processes when subjects work with textual representations</td>
</tr>
<tr>
<td>Somervell et al. [12]</td>
<td>Combination of graphical and textual representations were more efficient</td>
<td>Guidelines on the use of a combination of textual and graphical representations to improve subjects’ efficiency</td>
</tr>
<tr>
<td>Razali et al. [13]</td>
<td>Graphical formal specification vs. a purely textual formal specification</td>
<td>Combination of semi-formal and formal notations improves the subjects’ accuracy</td>
</tr>
<tr>
<td>Heijstek [2] et al.</td>
<td>Graphical and textual notations for software architecture</td>
<td>No difference in accuracy, more experienced subjects mostly preferred a textual representation</td>
</tr>
</tbody>
</table>


TROPOS

TROPOS is a goal-oriented requirements modeling approach based on concepts such as:

- **Actor** - typically representing a domain stakeholder
- **Goal** - representing a state of affairs desired by the actor
- **Task** - representing set of activities which operationalizes goals
- **Resource** - which is an element (such as information, device, database, …) whose presence is needed to support the satisfaction of goals or the execution of a task

And relationships such as:

- **AND/OR decomposition** of goals and tasks into sub-goals and sub-tasks
- **Means-ends** to describe the relationship between a goal and the task that fulfill it

Each concept or relationship has a visual counterpart
TROPOS: goal diagrams

Actor
Goal
Task
Resource
Means-ends relationship

Nurse

And decomposition of goals

monitor the stairs of the residence

history of the patient

biological sensors

monitor the rooms of the residence

patient log books

fall sensors

cameras

movement sensors

doors sensors

monitor patients movements
Our Empirical Study

Our Goal: Design and perform an experiment to investigate the impact of requirement representation on comprehension accuracy, time and strategy.

High Level Research Question: Does the document representation impact time or accuracy in program understanding tasks?

Perspective:
- Developers
- Researcher
Detailed Research Questions

- **RQ1**: Does the type of requirement representations (graphical vs. textual) impact the developers' **effort**, **time**, and **answer accuracy** in requirements comprehension tasks?

- **RQ2**: Does the structure of the representations lead developers to use **specific task-solving strategies** (top-down vs. bottom-up) during requirements comprehension tasks?

- **RQ3**: Given a graphical and textual representation of a requirements comprehension task, **is there any preferred representation** by the subjects?
Detailed Research Questions

- **RQ2**: Does the structure of the representations lead developers to use specific task-solving strategies (top-down vs. bottom-up) during requirements comprehension tasks?
RQ3: Given a graphical and textual representation of a requirements comprehension task, is there any preferred representation by the subjects?
## Experiment Design

<table>
<thead>
<tr>
<th>Goal</th>
<th>Study the impact of requirement representation</th>
</tr>
</thead>
</table>
| **Independent variables** | 1. Document representation  
          a) Graphical, b) Textual, c) Both; |
| **Dependent variables** | 1. Accuracy  
          2. Required time  
          3. Effort - Visual Effort |
| **Mitigating variables** | 1. Study level  
          2. English language proficiency  
          3. Mother language  
          4. Gender: male (M) or female (F) |

### Subjects’ Demography

<table>
<thead>
<tr>
<th>Academic background</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ph.D.</td>
<td>M.Sc.</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
</tr>
</tbody>
</table>
# Experiment Design

## Subjects’ Demography

<table>
<thead>
<tr>
<th>Academic background</th>
<th>Gender</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ph.D.</td>
<td>Male</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>M.Sc.</td>
<td>Female</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>B.Sc.</td>
<td>Male</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. What is/are the resource(s) that helps in having health emergency monitored?
   a) biological sensors
   b) biological sensors and patient log book

Goal: “patients’ falls monitored” that is operationalised by
   Task: “monitor the stairs” using
   Resource: “falls sensors” and “cameras”

Goal: “health emergencies monitored” that is operationalised by
   Task: “monitor history of the patient” via
   Resource: “patient log books”
   Task: “monitor health factors”
   Resource: “biological sensors”
Experiment Design

1. What is/are the resource(s) that helps in having health emergency monitored?
   a) biological sensors
   b) biological sensors and patient log book

- patients’ falls monitored
- monitor the stairs
- cameras
- falls sensors

- health emergencies monitored
- monitor health factors
- patient log books
- biological sensors

Helper
Actor
Goal
Task
Resource

Relevant
Experiment Design

- FaceLAB
  - Video-based
  - Two camera
  - One infrared

- Non-intrusive
  - No goggles
  - No wires
  - No sensing device
Result and Analysis: Visual Effort

- Visual effort
  - Calculated from eye-tracking data.
  - Calculated based on the amount of visual attention
    - less attention $\rightarrow$ less time $\rightarrow$ less effort

- Visual attention triggers the mental processes

- Two types of eye gaze data
  - Fixation
  - Saccade

- We use fixation to calculate effort
Result and Analysis: Visual Effort

- Convex hull: the smallest convex sets of fixations that contains all of a subject’s fixations*
  - Measure Average Fixation Duration (AFD) via convex hull as effort proxy
    - Smaller convex hull ➙ close fixations ➙ less effort

Result and Analysis: Visual Effort

Model
Actor: Nurse
Goal: “emergency detection” is AND decomposed in the:
  Task: “monitor the social residence doors” using
    Resource: “doors sensors”
  Task: “monitor patients movements” using
    Resource: “movement sensors”
Goal: “patients falls monitored” that is OR operationalised by
  Task: “monitor patients movements” using
    Resource: “movement sensors”
  Task: “monitor the stairs of the residence” using
    Resource: “falls sensors” and “cameras”
  Task: “monitor the rooms of the residence” using
    Resource: “falls sensors”

Goal: “health emergencies monitored” that is operationalised by
  Task: “monitor history of the patient” via
    Resource: “patient log books”
  Task: “monitor health parameters”
    Resource: “biological sensors”

Numbers of eye fixations

\[
FC(Q) = \sum_{a \in \text{tasks, all answers}} f(a)
\]  

\[
FR(\text{correct}) = \frac{\sum_{a \in \text{correct answer}} f(a)}{\sum_{a \in \text{correct answer} \cup \text{distracters}} f(a)}
\]

\[
FR(\text{distracters}) = \frac{\sum_{a \in \text{distracters}} f(a)}{\sum_{a \in \text{correct answer} \cup \text{distracters}} f(a)}
\]
Result and Analysis: RQ1

<table>
<thead>
<tr>
<th></th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
</tr>
<tr>
<td>Graphical</td>
<td>97%</td>
</tr>
<tr>
<td>Textual</td>
<td>98%</td>
</tr>
<tr>
<td>Mixed</td>
<td>96%</td>
</tr>
</tbody>
</table>

There is a significant difference in Time, but No significant difference in Accuracy.
Result and Analysis: RQ1

- Different model imply different areas of focus

The percentage of time that our subjects spent on different AOIs for Graphical model.

The percentage of time that our subjects spent on different AOIs for Textual model.
Result and Analysis: RQ1

- There is a significant difference in visual effort though Cohen-d is from medium up
- AFD(Q) -- Average Fixation Duration -- is borderline 0.07!
Result and Analysis: RQ2

1. What is/are the resource(s) that helps in having health emergency monitored?
   a) biological sensors
   b) biological sensors and patient log book

Goal area

Task area

Resource area

Relevant
Result and Analysis: RQ2

- The structure of our document makes subject use a top-down (goals to resources) or bottom up (resources to goals) strategy.

---

ECG for top-down strategy

ECG for bottom-up strategy
Result and Analysis: RQ3

- Subject Prefer Graphic Notation
Threats to validity

- Internal validity
  - Random ordering of stimuli
  - Provide comfortable environment

- External validity (generalisation of the results)
  - Students as subjects
  - “Only” 28 subjects
Conclusion

- Requirement representation has an impact

- Language distance has no impact
  - Closer to English is “better”

- Graphical representation is preferred but requires greater effort

- Gender has not impact
Conclusion

- Design and perform an eye-tracking experiment
- Investigate the impact of document representation on program comprehension
- Examine accuracy, time, effort, and preference