WP5 Evaluation

D54-D55 Evaluation Results V2 (V3)

Peter van Rosmalen

Contributors: all partners
WP5 Overview

• Introduction
• Evaluation results v2
• Evaluation results v3
• Overall conclusions
Objectives

- To evaluate the **effectiveness** of the idSpace system as a tool for **supporting in a context-aware manner** the creation of new ideas.
- To evaluate the **effectiveness** of the idSpace system as a tool for **representation, storage, and management of ideas**.
- To assess the **usability** of the idSpace system’s interfaces
- To assess the **user’s experience** with the interaction
- To identify any **specific problems** with the idSpace system
## Main Activities

<table>
<thead>
<tr>
<th>Last review - January</th>
<th>Evaluation Definition &amp; Set-up &amp; Preparation</th>
</tr>
</thead>
<tbody>
<tr>
<td>January - 6 February</td>
<td>1. User testing:</td>
</tr>
<tr>
<td></td>
<td>MORPH; SAS; LiNK; OUNL, AAU, UNI Hildesheim; UPRC</td>
</tr>
<tr>
<td></td>
<td>2. Test claim validation (SAS)</td>
</tr>
<tr>
<td></td>
<td>3. Heuristic evaluation (SAS)</td>
</tr>
<tr>
<td></td>
<td>&amp; Input to WP4 V3</td>
</tr>
</tbody>
</table>

| February               | Focused evaluations:                          |
|                       | 1. CA (Context Awareness) - Component:      |
|                       | recommendations portlets                     |
|                       | 2. Reasoning portlet                         |
|                       | & Input to WP4 v3                            |

| March (April)          | Expert evaluation: v3                        |
|                       | D54-55 Final version                         |
|                       | & Input to WP4 v3                            |
Evaluation Results V2
V2 Evaluation Perspective

- careful,
- rigorous,
- *eager to learn*,
- (fairly) large-scale.
Results - 1. User Testing

- 6 test groups:
  - Fixed set-up & materials (minor local adaptations):
    - Briefing package
    - General idSpace and tool intro
    - Fixed set of tasks
    - background questionnaire; post questionnaire; observers
  - 24 users; 6 moderators
  - 4 ‘closed’, ‘local’ groups; 2 ‘open’, ‘distance’ groups
  - Variety of initial problems statements
Results - 1. User Testing

User appreciated:

- the access to the system via the web - without installation requirements
- the structured, distributed and collaborative approach to ideation

However, ..........
### Results - 1. User Testing

<table>
<thead>
<tr>
<th>Criteria/ scores</th>
<th>SAS</th>
<th>UPRC</th>
<th>UNI Mobile</th>
<th>LiNK</th>
<th>MORPH</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>effectiveness</td>
<td>2.33</td>
<td>2.39</td>
<td>2.85</td>
<td>2.72</td>
<td>3.29</td>
<td>2.72</td>
</tr>
<tr>
<td>learnability</td>
<td>1.97</td>
<td>1.75</td>
<td>2.07</td>
<td>2.56</td>
<td>2.88</td>
<td>2.25</td>
</tr>
<tr>
<td>user satisfaction</td>
<td>1.77</td>
<td>2.14</td>
<td>2.70</td>
<td>2.61</td>
<td>3.00</td>
<td>2.44</td>
</tr>
<tr>
<td>efficiency</td>
<td>2.29</td>
<td>2.21</td>
<td>2.60</td>
<td>2.82</td>
<td>3.14</td>
<td>2.61</td>
</tr>
<tr>
<td>Average</td>
<td>2.09</td>
<td>2.12</td>
<td>2.56</td>
<td>2.68</td>
<td>3.08</td>
<td></td>
</tr>
</tbody>
</table>

* scores aligned, so 1 is very negative, 2 negative, 3 neutral, 4 positive and 5 very positive
Results - 1. User Testing

- - - extensive list of errors, problems with task completion, functional and usability issues, lack of background data (see the deliverable for details)

+ still, users do mention positive points and do express their interest (in particular the 'open' 'distance' groups*: UNI-Mobile & MORPH-TMRA)

* characteristics: multi-organisation; only option: distance collaboration; research/university background.
Results - 2. Test Claim validation

A system walkthrough by 2 SAS evaluators verifying 23 test claims based on 50* (of 89) ‘implemented’ user requirements (D5.1):
- 23 (46%) compliant
- 12 (24%) partially compliant
- 15 (30%) not compliant (4 of which being core requirements)

An analysis of the SAS user evaluation results on 5 usability requirements:
- 3 (not compliant)
- 2 (partially compliant)

* note: some of the requirements, e.g. UR5.1, UR5.2, may be unnecessarily ambitious because of existing, good alternatives such as chat, forum, email, Skype. In line with this e.g. for idSpace v3 the chat-portlet has been removed.
Results - 3. Expert Evaluation

Method:
- 3 experts, a system test following the user tasks
- Briefing - evaluation - debriefing
- Focus: identify issues & there severity based on a set of accepted heuristics

Heuristics:
- (10) Usability (Nielsen, 1994)
- (3) Creativity Support Tools (Warr, 2007)
- (5) Computer Supported Collaborative Creativity (Herrmann, 2009)
- (1) Provide Protection

Combined severity assessment:
- Frequency: common or rare.
- Impact: difficult or easy to overcome.
- Persistence: a one-time problem or would users repeatedly be bothered

Resulting in 5 point scale: 0 (no problem) to 4 (“catastrophe”, i.e. imperative to fix)
## Results - 3. Expert Evaluation

<table>
<thead>
<tr>
<th>Heuristics</th>
<th>Number</th>
<th>Severity 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability</td>
<td>61</td>
<td>7</td>
</tr>
<tr>
<td>Creativity Support Tools</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Computer Supported Collaborative Creativity</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Provide Protection</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>
**V2 Evaluation Conclusions**

**Perspective:**
- Careful, rigorous, *eager to learn*, (fairly) large-scale evaluation

**Important lesson learnt:**
- .. leaving out a version: removes an important feedback loop & lessens the integrative perspective.....

**Results:**
- **Software:** serious improvements required ('more by less'):
  - Focus on main problems identified, in particular:
    - Usability
    - Assure that the CA (Context Awareness) component, reasoning portlets etc resulting from the Research WPs do work
- Improve documentation
- Moderator role requires serious expertise (experience) both in ideation but also in online computer supported collaborative teamwork

**Additional work:**
- Extend the evaluation with an initial v3 trial
Evaluation Results “V3”
Expert User Evaluation V3

Objective
  • To evaluate and check the v3 improvements

Set-up. Similar as user test v2:
  • 2 expert user (completed previously at least 4 sessions)
  • Fixed set of tasks
  • 'Familiar' problem statement (mobile learning)

Results:
  • Many usability issues addressed (approximately 80-90% of the level 4 and 3 issues and a number of the “easy” ones)
  • Screens redesigned: core vs supportive portlets
  • Ideation screen redesigned
  • CA-component & reasoning portlet corrected & improved
V3 Examples

• Navigation
V3 Examples

- Navigation
- Current selection & status
V3 Examples

• Navigation
• Current selection & status
• Resources added
V3 Examples

- Navigation
- Current selection & status
- Resources added
- Mouseover text
V3 Examples

- Navigation
- Current selection & status
- Resources added
- Mouseover text
- Ideation screen redesign
On the move: Mobile learning for working professionals
Reasoning Portlet Evaluation V3

Objective:
- evaluate the effectiveness of the six transformation rules.

Set-up:
- A replicated experiment: 20 student participants in 2 groups:
- An idea map, examples for each rules + one random example
- Rate Association: (1) correct, (2) helpful for new ideas, and (3) supportive to identifying innovative ideas

Results:
- Results are mixed, as expected, no absolute successful rule:
  - Correctness: rule 4, 5 and 6
  - Helpfulness: rule 3, 4 and 6
  - Supportiveness: rule 6
Reasoning Portlet Evaluation V3

Objective:
- evaluate the effectiveness of the six transformation rules.

Set-up:
- A replicated experiment: 20 student participants in 2 groups:
- An idea map, examples for each rules + one random example
- Rate Association: (1) correct, (2) helpful for new ideas, and (3) supportive to identifying innovative ideas

Results:
- Results are mixed: no absolute successful rule
- Knowledge gained: resulting in recommendation for next version: remove rule 1, keep rule 2-5, boost rule 6
- ........... In addition: the pure act of prompting may contribute by activating and broadening the user’s thinking
Context Awareness Evaluation V3

Objective:
- to evaluate the CA component’s (recommendation of user, idea, solution) usefulness and usability.

Set-up:
- Context relevant data pre-entered
- 11 post graduate students
- 3 tasks, i.e. retrieve and validate recommendations on:
  (1) solutions to related problems
  (2) suitable users to add to your team
  (3) ideas that may be useful/relevant for solving the particular problem.
Context Awareness Evaluation V3

Objective:
- to evaluate the CA component’s (recommendation of user, idea, solution) usefulness and usability.

Results:
(1) Solutions to related problems - useful & helpful (mean 5.14*):
   in particular how to formulate a possible statement/description
(2) Suitable users - useful & convenient (mean 5.11):
   in particular as the number of potential users increases
(3) Ideas - good (mean 4.86)

Overall usefulness & ‘ease of use’: mean approximately 5

* 7 point Likert scale: 1=lowest 7 highest
Evaluation Conclusions V3

The three studies indicate (*be it not conclusive because of their size*) that:

- many of the issues raised in the v2 evaluation have been addressed (approximately 80-90% of the level 4 and 3 issues and a number of the “easy” ones)

- the *usability* of the idSpace system’s interfaces &
- the *user’s experience* with the interaction has been improved.

Moreover, in particular the possibility to add external resources and the results of the recommender and reasoning portlets do positively influence the *effectiveness* of the idSpace system.
Overall Conclusions
A rigorous evaluation of idSpace v2 indicating:

- the user appreciated the access to the system via the web - without installation requirements - and the structured, distributed and collaborative approach to ideation.
- many small and big issues (in six user sessions, an heuristic evaluation & a test claim validation) having a strong negative impact on the overall appreciation of the V2 release.
- each of the four major indicators of user experience, i.e., effectiveness, efficiency, learnability and user satisfaction scored negative.

idSpace v2

“I can feel that there is a potential for a useful tool to be developed to help ideation”

“The prototype nicely presents the core idea of the project. But there are many usability issues in the software and some logical problems in the ideation and solution screens”
Overall Conclusions

An additional evaluation study of v3 (an expert user evaluation & two targeted studies indicating that:

- many (80-90%) of the issues raised have been addressed
- usability, user’s experience and effectiveness of the idSpace system have been improved.
Overall Conclusions

the idSpace system v3 offers:
• access to the system via the web – without installation requirements
• a structured, guided, distributed and collaborative approach to ideation
• effective support made available through CA & reasoning portlets
Discussion & Questions