Recovering Interaction
Design Patterns in Web
Applications

P. Tramontana
A.R. Fasolino
Dipartimento di Informatica e Sistemistica
University of Naples Federico II, Italy

G.A. Di Lucca
RCOST – Research Centre on Software Technology
University of Sannio, Benevento, Italy
Web Interaction Design Patterns

- The design of User Interface is a critical phase of Web Applications development
  - Many quality attributes of Web Applications depend on the interface (Usability, Accessibility, …)
- Use of interaction patterns speed up design and implementation of high quality Web Applications
- Identification of Web Interaction Design Patterns helps us for the comprehension, maintenance, reengineering, reuse of Web Application
Some catalogues of Web Interaction Design Patterns have been proposed

Martin van Welie in website [www.welie.com](http://www.welie.com) reports a list comprehending many different patterns

An abstract description of each pattern is provided by natural language ...
  - not enough to support automatic identification!
An example of WIDP Description: the Login pattern

**Problem** The users need to identify themselves so that stored data about/of them can be used in the process they are in.

**Context** When users frequently return to a site that uses large amounts of data about or belonging to the users, it is convenient to have users enter that information once and use it again for future visits to the site. Usually the information that is stored is personal information and can include name, age, gender, shipping addresses, stock portfolio, bank account numbers and credit card numbers. In order to be able to access their data, users must complete their Registration first.

For many site types it can be convenient to store information of/about visitor. Often these are E-commerce Site, Community Site or Web-based Application such as electronic banking applications.

**Solution** When needed, ask the users to login using a combination of a username and a password
Identification of WIDPs

Examples of Login pattern

What are the common characteristics of the login pattern implementations?

- a Form
- a text field
- a password field
- a submit button
- word ‘password’
- ...
User Interface Model

- Clues useful to support the identification of WIDPs may be recovered analyzing HTML structures and textual part of Client Pages.
Features

- **Structural features**
  - a table with 2 or more rows in the page
  - 2 or more anchors in the page
  - a form in the page
  - a password field in the page

- **Lexical features**
  - any of the words ‘login’, ‘username’ (a login synonym), ‘nome utente’ (an Italian translation for login or username) in the page

What are the most characteristics features for a given WIDP?
Characterization of Patterns’ features

- Given a pattern a feature is **characteristic** one if:
  - it is “*frequently*” retrieved in the implementation of the pattern and
  - it isn’t “*frequently*” retrieved in the implementation of any other pattern
Measure of the Characterization

- Given a set of implementations of each pattern (training set) we define:
  - Frequency of a Feature F for a Pattern P
    \[ Freq(P, F) = \sum_{wp \in \text{TrainingSet}(P)} \frac{Occ(wp, F)}{\text{Card}(\text{TrainingSet}(P))} \]
  - Specificity of a Feature F for a Pattern P
    \[ Spec(P, F) = \begin{cases} 
    Freq(P, F) - \text{Average}(F) & \text{if } Freq(P, F) > \text{Average}(F) \\
    0 & \text{Elsewhere} 
    \end{cases} \]
  - Characterization Weight of a Feature F for a Pattern P
    \[ CW(P, F) = Freq(P, F) \times Spec(P, F) \]
A Three-phase approach to identify Web Interaction Design Patterns in Web Pages

The identification of patterns is based on the analysis of the occurrences of the characteristic features present in the pages.
Training Phase

- A training set of Web Pages actually containing samples of each pattern have to be selected.

Training Web Pages:
  - May be written using different human languages (e.g. English and Italian)
  - May belong to different application domains
  - May be developed with different programming styles (e.g. automatically generated or manually developed)

- For each couple (pattern, feature) the Characterization Weight value is evaluated
Candidature phase

- for each pattern and for each Web Page, an index of the probability that a pattern P is contained in a Web Page WP is evaluated

\[
\text{Likelihood}(WP, P) = \frac{\sum_{f \in \text{FeatureSet}(P)} CW(P, f) * \text{Occ}(WP, f)}{\sum_{f \in \text{FeatureSet}(P)} CW(P, f)}
\]

- We assume that a pattern is contained in a Web Page if the Likelihood value is greater than a fixed threshold

- Controlled experiments have been carried out to evaluate the threshold value maximizing the number of correct candidatures
Validation phase

- The correctness of the (Web Page, Pattern) couples proposed during the Candidature phase must be assessed by an expert that verifies if the Web Page actually contains the proposed Pattern.

- Web Pages containing validated patterns may be added to the training set of the pattern.
Architecture of the system

- Web pages
- Reverse Engineering Tool
- Pattern Repository
- WA Information Repository
- Likelihood Evaluator
- Pattern Identifier
- Pattern Validator
- CW Tuner
- CW Evaluator
- Validated patterns
- Training Set
- Reverse Engineering Tool
Experiments have been carried out:
- To validate the approach
- To tune the value of the decision threshold
- To measure the degree of correctness of the patterns automatically identified by the system

Experiments involved:
- 6 Web Interaction Design Patterns (Guestbook, Login, Poll, Registration, Search, Sitemap)
- 216 Features (180 structural features and 36 lexical features)
## Pattern List

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guestbook</td>
<td>a view of the list of messages written by the visitors of a website</td>
</tr>
<tr>
<td>Login</td>
<td>an authentication module for inserting personal identification information needed to access private services</td>
</tr>
<tr>
<td>Poll</td>
<td>a module to insert a vote for a poll</td>
</tr>
<tr>
<td>Registration</td>
<td>a module to insert personal data needed to register to a service</td>
</tr>
<tr>
<td>Search</td>
<td>a module to insert keywords for a search on a search engine</td>
</tr>
<tr>
<td>Sitemap</td>
<td>a view of the map of the pages of a website</td>
</tr>
</tbody>
</table>
**Guestbook**

**Tuesday, 28. September 2004 17:51**

Thank you for stopping by my site. Here you can leave your mark.

Total Records: **4394**  Records Viewed Per Page: **20**

<table>
<thead>
<tr>
<th>Name</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>4394</td>
<td>Tuesday, 28. September 2004 17:48</td>
</tr>
<tr>
<td>jgvries</td>
<td>Anyone can tell me where i can set the admin pass for the advanced guestbook?? :S</td>
</tr>
<tr>
<td>Location: stadskaanaal</td>
<td></td>
</tr>
<tr>
<td>4393</td>
<td>Tuesday, 28. September 2004 14:06</td>
</tr>
<tr>
<td>Ninja X</td>
<td>orjgolksnrheognedgdt</td>
</tr>
<tr>
<td>Location: -</td>
<td></td>
</tr>
<tr>
<td>4392</td>
<td>Tuesday, 28. September 2004 12:54</td>
</tr>
<tr>
<td>and688</td>
<td>cool site!</td>
</tr>
<tr>
<td>Location: and688(at)hotmail(dot)com</td>
<td></td>
</tr>
</tbody>
</table>
Login

Benvenuto in webmail.unina.it

Nome utente: [campo vuoto]
Password: [campo vuoto]
Lingua: [campo vuoto]

Inizio sessione
Poll

Poll

Sondaggio Spettacoli & Cultura

Il Festival di Bonolis

Il Festival di Sanremo targato Bonolis sta avendo un grande successo di ascolti. E voi che ne pensate?

1. Mi piace
2. Non mi piace
3. Non lo so

1.520 voti alla 20:04
Sondaggio aperto alle 11:36 del 03-03-2005

Risultati

AVVERTENZA
I sondaggi online di Repubblica.it non hanno, ovviamente, un valore statistico. Si tratta di rilevazioni aperte a tutti, non basate su un campione elaborato scientificamente. Hanno quindi l'unico scopo di permettere ai lettori di esprimere la propria opinione sui temi di attualità. Le percentuali non tengono conto dei valori decimali. In alcuni casi, quindi, la somma può risultare superiore a 100.
Registration

CSMR
21-23 March 2005
REGISTRATION FORM

To register, please complete this registration form and submit it using the button below.

Further details will be sent with a receipt and confirmation of booking.

Fields marked with * must be completed.

*Title: 
*First Name: 
*Last Name: 
This information will be used in the production of badges.

*Institution: 
*Address: 

Evaluation of the Characterization Weight

- The Characterization Weight has been evaluated analyzing a Training set comprehending 108 Web Pages.
- Each Web Page of the Training Set contains only one of the considered pattern.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guestbook</td>
<td>15</td>
</tr>
<tr>
<td>Login</td>
<td>25</td>
</tr>
<tr>
<td>Poll</td>
<td>13</td>
</tr>
<tr>
<td>Registration</td>
<td>14</td>
</tr>
<tr>
<td>Search</td>
<td>20</td>
</tr>
<tr>
<td>Sitemap</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Best Poll Features</th>
<th>CW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word ‘poll’</td>
<td>0.5644125</td>
</tr>
<tr>
<td>2 or more select buttons in a form</td>
<td>0.4741561</td>
</tr>
<tr>
<td>Word ‘vote’</td>
<td>0.4399831</td>
</tr>
<tr>
<td>Word ‘results’</td>
<td>0.4216681</td>
</tr>
<tr>
<td>2 or more radio button in a table</td>
<td>0.3189687</td>
</tr>
<tr>
<td>1 radio button in a cell</td>
<td>0.3031277</td>
</tr>
</tbody>
</table>
Recall and Precision

To assess the correctness of the results of the approach, the Recall and Precision measures have been defined:

- **Recall**: Number of correct candidate couples (web page, pattern) / Number of couples (web page, pattern) to identify

- **Precision**: Number of correct candidate couples (web page, pattern) / Number of candidate couples (web page, pattern)
A preliminary test has been carried out to assess the effectiveness approach.

The preliminary identification test involves the same Web Pages constituting the Training Set.

Recall and Precision values have been evaluated, varying threshold value.
## Preliminary Identification Test

<table>
<thead>
<tr>
<th>Threshold 0.8</th>
<th>0.7</th>
<th>0.6</th>
<th>0.5</th>
<th>0.4</th>
<th>0.3</th>
<th>0.2</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>43%</td>
<td>64%</td>
<td>82%</td>
<td>86%</td>
<td>94%</td>
<td>98%</td>
<td>99%</td>
</tr>
<tr>
<td>Precision</td>
<td>46/49</td>
<td>69/75</td>
<td>89/112</td>
<td>93/135</td>
<td>101/174</td>
<td>106/234</td>
<td>107/325</td>
</tr>
<tr>
<td>%</td>
<td>94%</td>
<td>92%</td>
<td>79%</td>
<td>69%</td>
<td>58%</td>
<td>45%</td>
<td>33%</td>
</tr>
</tbody>
</table>

- A good trade-off between Recall and Precision has been reached for Threshold=0.6
- False positives and true negatives are due to the extreme similarity between some implementations of the patterns
Another experiment has been carried out, involving a test set of 108 Client Pages that have been extracted from the WWW.

Each of the Web Pages belonging to the test set contains zero, one or more patterns.

An expert software engineering stated which and how many patterns were included in the set of pages.
### Identification Test

<table>
<thead>
<tr>
<th>Threshold (≥ 0,8)</th>
<th>Recall 20/55</th>
<th>Recall 31/55</th>
<th>Recall 44/55</th>
<th>Recall 45/55</th>
<th>Recall 51/55</th>
<th>Recall 52/55</th>
<th>Recall 54/55</th>
<th>Recall 54/55</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,7</td>
<td>36%</td>
<td>56%</td>
<td>80%</td>
<td>82%</td>
<td>93%</td>
<td>95%</td>
<td>98%</td>
<td>98%</td>
<td>83%</td>
</tr>
<tr>
<td>0,6</td>
<td>31/55</td>
<td>44/55</td>
<td>45/55</td>
<td>51/55</td>
<td>52/55</td>
<td>54/55</td>
<td>54/55</td>
<td>% 83%</td>
<td></td>
</tr>
<tr>
<td>0,5</td>
<td>44/55</td>
<td>45/55</td>
<td>51/55</td>
<td>52/55</td>
<td>54/55</td>
<td>54/55</td>
<td>54/55</td>
<td>% 83%</td>
<td></td>
</tr>
<tr>
<td>0,4</td>
<td>45/55</td>
<td>51/55</td>
<td>52/55</td>
<td>54/55</td>
<td>54/55</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td></td>
</tr>
<tr>
<td>0,3</td>
<td>51/55</td>
<td>52/55</td>
<td>54/55</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td></td>
</tr>
<tr>
<td>0,2</td>
<td>52/55</td>
<td>54/55</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td></td>
</tr>
<tr>
<td>0,1</td>
<td>54/55</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td>% 83%</td>
<td></td>
</tr>
</tbody>
</table>

- A good trade-off between Recall and Precision has been reached (again) for Threshold=0.6
- The performance are worst than in preliminary test but are quite acceptable
Conclusions

- Automatic identification of Web Interaction Pattern is not a simple task:
  - Interaction functionality may be implemented in much different ways
- An approach has been proposed to identify Patterns implemented in Web Pages on the basis of the identification of the common features characterizing the pattern
- Experiments that have been carried out shows the feasibility of the proposed approach
Future Works

- Further assessing experiments must be carried out to measure the correctness of the results of the approach by varying:
  - The number of patterns to identify
  - The number of samples in the training set
  - The set of features (e.g. adding features resulting from the combination of the basic features considered in the presented experiment)