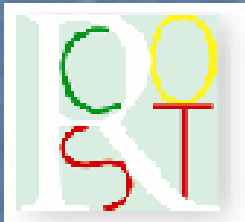


# Recovering Interaction Design Patterns in Web Applications

---



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 depends 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

# Web Interaction Design Pattern catalogues

---

- 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



Accesso a .NET Passport [Guida](#)

Indirizzo di posta elettronica

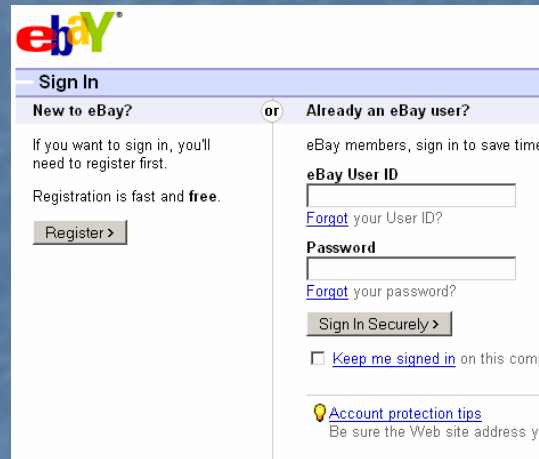
Password

Accedi automaticamente

Non memorizzare l'indirizzo di posta elettronica per gli accessi futuri (selezionare se si utilizza un computer pubblico).

Microsoft .net

Non hai un profilo .NET Passport? [Registrati ora!](#)



Sign In

New to eBay?  or Already an eBay user?

If you want to sign in, you'll need to register first.

Registration is fast and free.

eBay members, sign in to save time

eBay User ID

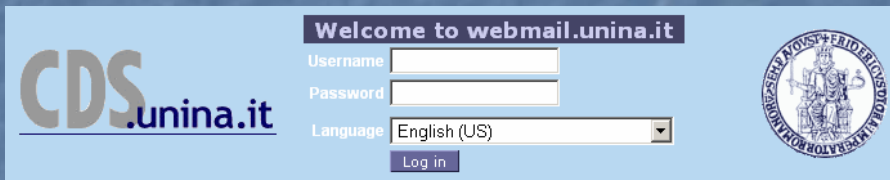
[Forgot](#) your User ID?

Password

[Forgot](#) your password?

[Keep me signed in](#) on this computer

 [Account protection tips](#)  
Be sure the Web site address y



Welcome to webmail.unina.it

CDS.unina.it

Username

Password

Language

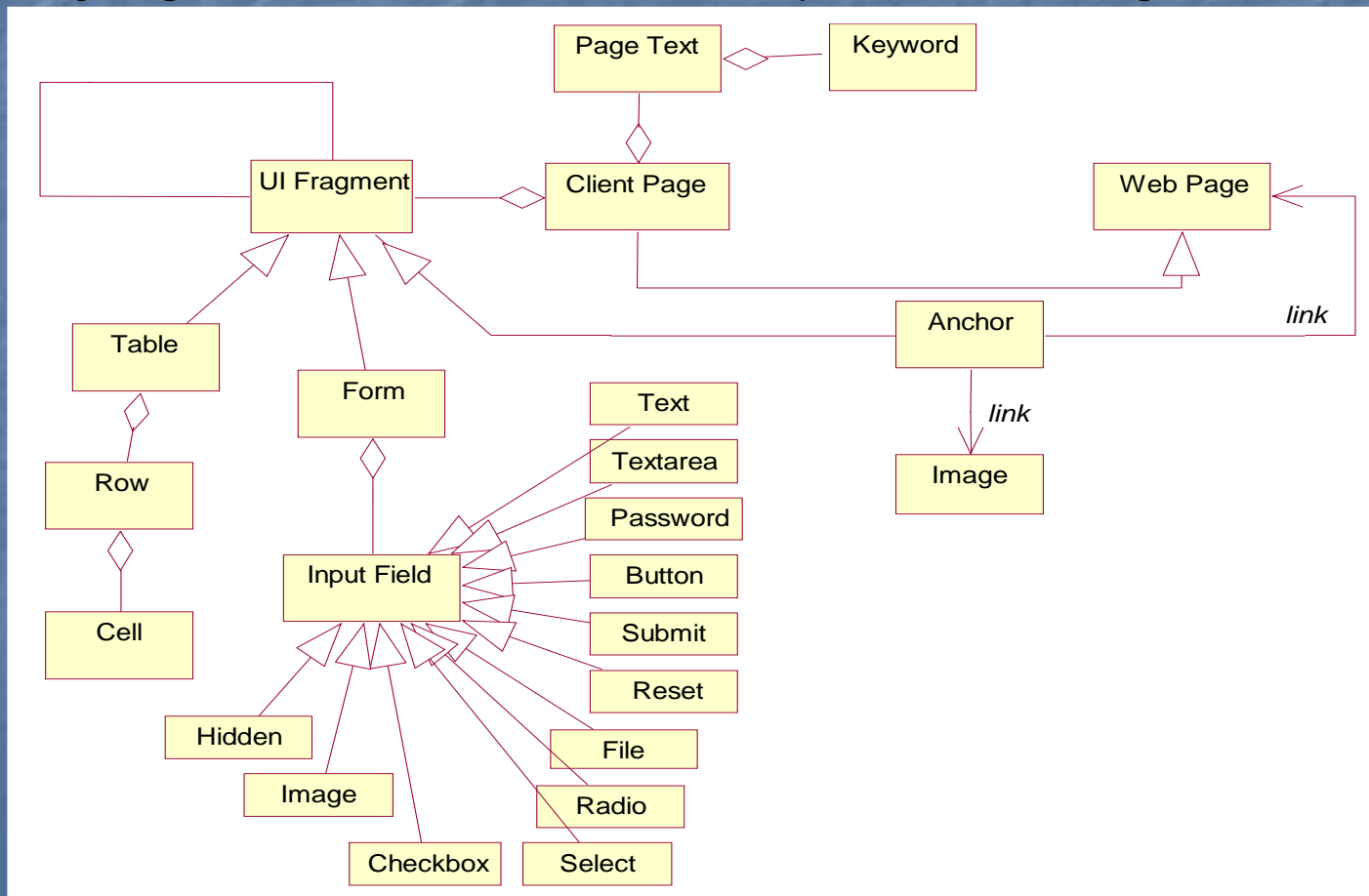


What are the common characteristics of the login pattern implementations?

- 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

$$Freq(P, F) = \sum_{wp \in TrainingSet(P)} \frac{Occ(wp, F)}{Card(TrainingSet(P))}$$

- Specificity of a Feature F for a Pattern P

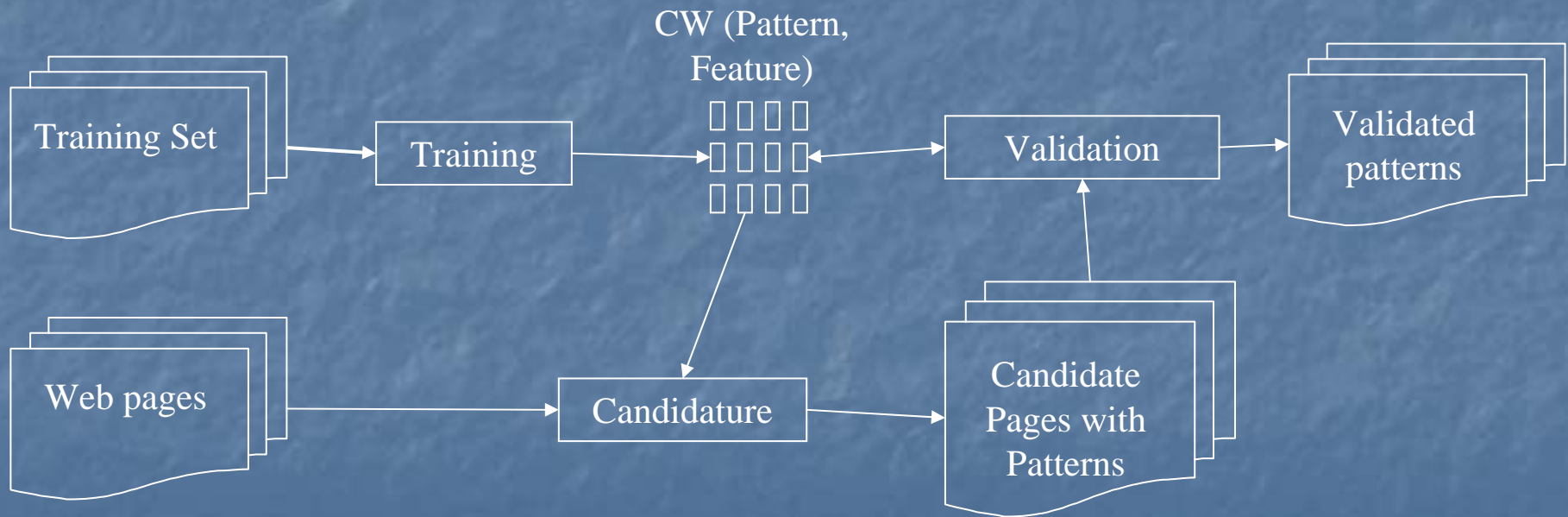
$$Spec(P, F) = \begin{cases} Freq(P, F) - Average(F) & \text{if } Freq(P, F) > Average(F) \\ 0 & \text{Elsewhere} \end{cases}$$

- Characterization Weight of a Feature F for a Pattern P

$$CW(P, F) = Freq(P, F) * Spec(P, F)$$

# 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

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

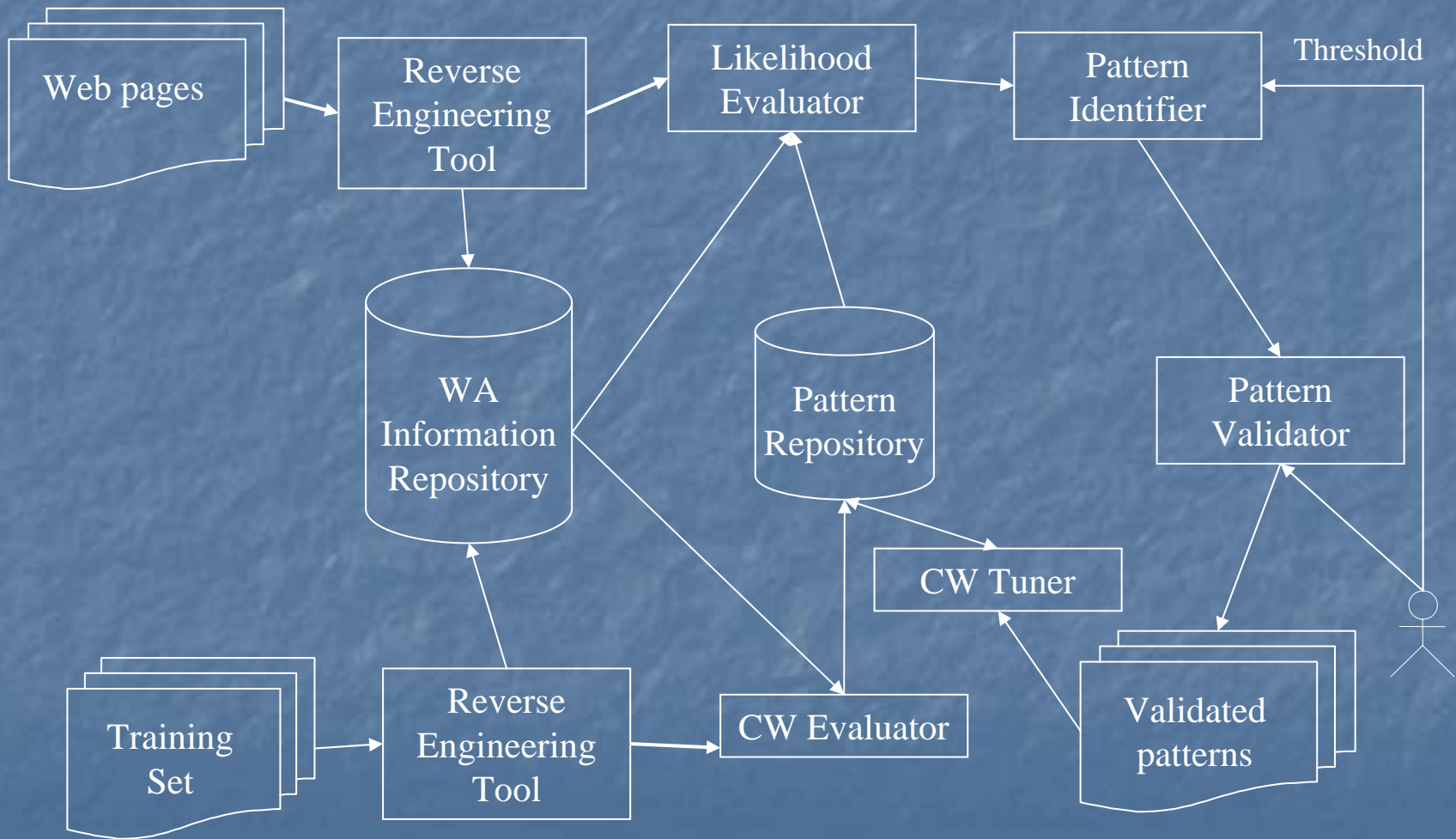
- We assume that a pattern is contained in a Web Page if the Likelihood value is greater than a fixed threshold
  - Controlled experiments have been carried out to evaluate the threshold value maximizing the number of correct candidatures

# 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









# Experimentation

---

- 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

<h2>Guestbook</h2>	<p>a view of the list of messages written by the visitors of a website</p>	
<h2>Login</h2>	<p>an authentication module for inserting personal identification information needed to access to private services</p>	
<h2>Poll</h2>	<p>a module to insert a vote for a poll</p>	
<h2>Registration</h2>	<p>a module to insert personal data needed to register to a service</p>	
<h2>Search</h2>	<p>a module to insert keywords for a search on a search engine</p>	
<h2>Sitemap</h2>	<p>a view of the map of the pages of a website</p>	



# Guestbook

## Guestbook

[Sign the Guestbook](#) | [Administration](#)

Tuesday, 28. September 2004 17:51

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

Guestbook

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

[Next Page](#)

Name	Comments
<b>4394)</b> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <b>jgvriez</b> <input type="checkbox"/> jgvriez(at)gmail(dot)com Location: stadskanaal	Tuesday, 28. September 2004 17:48 <input type="checkbox"/> <input type="checkbox"/> <hr/> Anyone can tell me where i can set the admin pass for the advanced guestbook??? :S
<b>4393)</b> <input type="checkbox"/> <input type="checkbox"/> <b>Ninja X</b> <input type="checkbox"/> Location: -	Tuesday, 28. September 2004 14:06 <input type="checkbox"/> <hr/> <input type="checkbox"/> orjgolksnrheogtine.dgdt
<b>4392)</b> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <b>and688</b> <input type="checkbox"/> and688(at)hotmail(dot)com Location:	Tuesday, 28. September 2004 12:54 <input type="checkbox"/> <input type="checkbox"/> <hr/> cool site!

# Login

**CDS**.unina.it

**Benvenuto in webmail.unina.it**

Nome  
utente

Password

Lingua

[Inizio sessione](#)



# Poll

la Repubblica.it

Cerca con Google



[Homepage](#) [Sport](#) [Economia](#) [Tecnologia & Scienze](#) [Spettacoli & Cultura](#) [Radio](#)

[Politica](#) [Cronaca](#) [Esteri](#) [Arte](#) [Motori](#) [Week-in](#) [Scuola & Università](#) [Lavoro](#) [Cronologia](#) [Meteo](#) [Foto](#) [Video](#)

ZOPPINI  
FIRENZE

LEGGI  
IL GIORNALE



MOBILE

via SMS

i-mode

Umts

Gprs

SERVIZI

Repubblica Voice

FreeInternet

Cerca nel sito

Scriveteci

Repubblica@scuola

SUPPLEMENTI

Dweb

Affari & Finanza

Salute

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 seguo

Vota

18520 voti alle 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



Ascolta  
Capital.it



Ascolta  
Guarda  
DeeJay Tv



Ascolta  
m2o.it

Kataweb



Lettera Finanziaria



ilPassaporto



Shopping online



Libreria online



Stampa le tue foto



Music Download



Cittadino lex



Tutto sui cellulari



Fantacalcio



Vetrina Immobiliare



Oroscopo

I segni, le  
previsioni, la  
posta di Horus

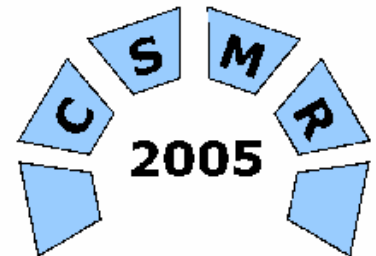
Trovacinema

Tutti i film, le  
sale, le  
recensioni

# 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:

# Search

---



**Web** [Images](#) [Groups](#) [News](#) [Froogle](#) [Local](#)<sup>New!</sup> [Desktop](#) [more »](#)

[Advanced Search](#)  
[Preferences](#)  
[Language Tools](#)

[Advertising Programs](#) - [About Google](#) - [Go to Google Italia](#)

©2005 Google - Searching 8,058,044,651 web pages

# Sitemap

Microsoft.com Home | Site Map

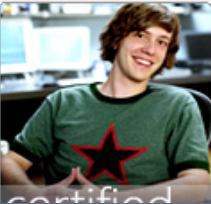
**Microsoft TechNet**

Search Microsoft.com for:

Search for

TechNet Home  
Products & Technologies  
IT Solutions  
Security  
Interop & Migration  
Desktop Deployment  
Script Center  
Community  
Downloads  
IT Training & Certification  
Troubleshooting & Support  
TechNet Program

Archive  
TechNet Site Map  
TechNet Worldwide



**TechNet Site Map**

- **Products and Technologies**  
[Products and Technologies: Overview](#)
  - [Server Operating Systems](#)
  - [Servers](#)
  - [Desktop Operating Systems](#)
  - [Desktop Products and Technologies](#)
  - [Embedded Operating Systems](#)
- **IT Solutions**  
[IT Solutions: Overview](#)
  - [Windows Server System Reference Architecture](#)
  - [Common IT Scenarios](#)
  - [Core IT Services](#)
  - [Mobility](#)
  - [Networking](#)
  - [Microsoft IT Showcase](#)
  - [Indexes](#)
- **Security**  
[Security Overview](#)
  - [Microsoft Security Bulletin Search](#)
  - [Virus Alerts](#)
  - [Security Tools](#)
  - [Product and Technology Security Centers](#)
- **Script Center**  
[Script Center: Overview](#)
  - [Script Repository](#)
  - [Solutions Center](#)
  - [Upcoming Scripting Webcasts](#)
  - [Tools and Utilities](#)
  - [Hey, Scripting Guy!](#)
  - [Tales from the Script](#)
- **Community**  
[Community: Overview](#)
  - [Newsgroups](#)
  - [Chats](#)
  - [Events & Webcasts](#)
  - [Columns](#)
  - [TechNet Radio](#)
- **Downloads**  
[Downloads: Overview](#)
- **IT Training and Certification**  
[IT Training and Certification: Overview](#)
  - [TechNet Virtual Lab](#)
- **Troubleshooting and Support**  
[Troubleshooting and Support: Overview](#)

# 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

Pattern	#
Guestbook	15
Login	25
Poll	13
Registration	14
Search	20
Sitemap	21

Best Poll Features	CW
Word 'poll'	0,5644125
2 or more select buttons in a form	0,4741561
Word 'vote'	0,4399831
Word 'results'	0,4216681
2 or more radio button in a table	0,3189687
1 radio button in a cell	0,3031277

# 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)*



# Preliminary Identification Test

---

- 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

Threshold	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1
Recall	46/108	69/108	89/108	93/108	101/108	106/108	107/108	108/108
%	43%	64%	82%	86%	94%	98%	99%	100%
Precision	46/49	69/75	89/112	93/135	101/174	106/234	107/325	108/487
%	94%	92%	79%	69%	58%	45%	33%	22%

- 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

# Identification Test

---

- 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

Threshold	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1
Recall	20/55	31/55	44/55	45/55	51/55	52/55	54/55	54/55
%	36%	56%	80%	82%	93%	95%	98%	98%
Precision	20/24	31/39	44/67	45/112	51/171	52/229	54/347	54/590
%	83%	79%	66%	40%	30%	23%	16%	9%

- 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)