Evaluation Methods for Web Application Clustering

P.Tonella, F.Ricca, E.Pianta, C.Girardi: ITC-Irst

G.DiLucca: Rcost, Università del Sannio

A.R.Fasolino, P.Tramontana: Università di Napoli

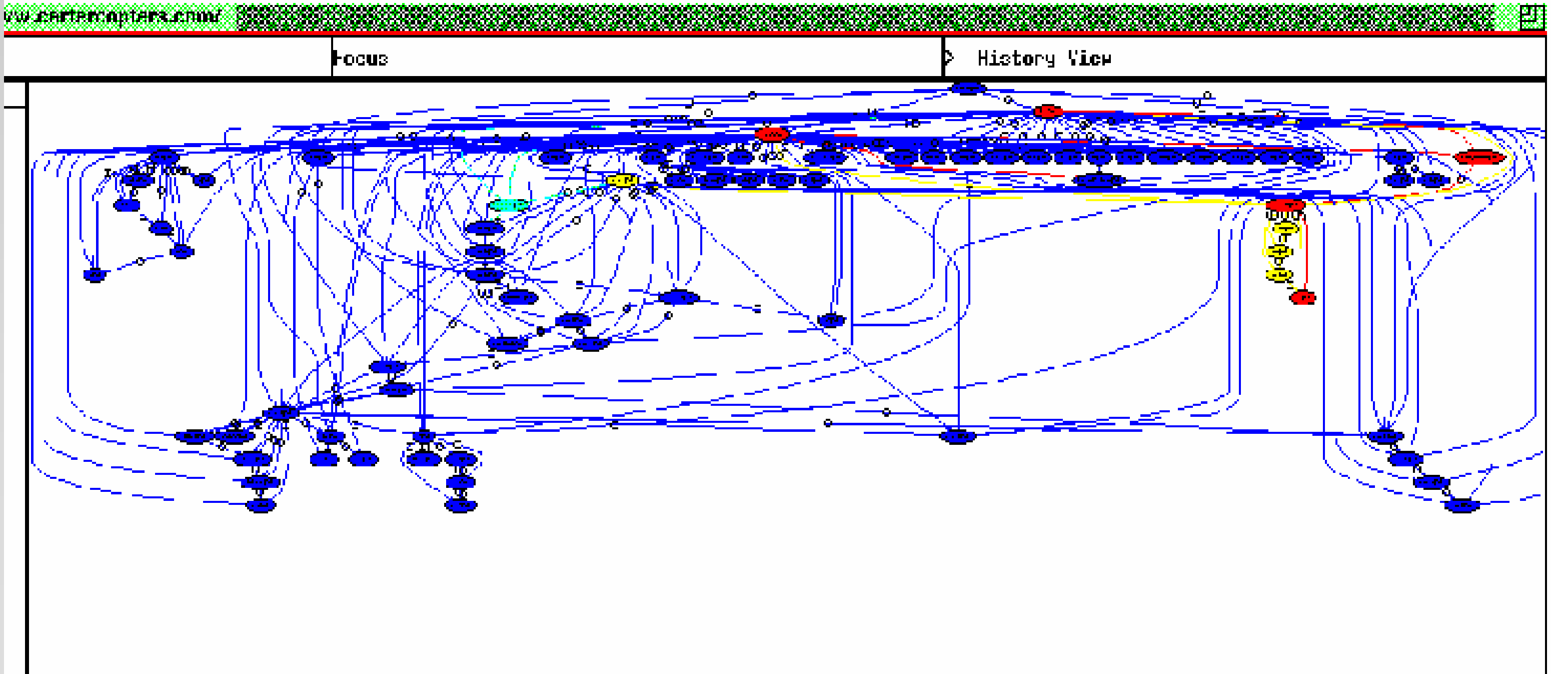
Web site understanding

Reverse engineering techniques have the potential **to support Web site understanding** by providing views that show the organization of a site.

Web pages = nodes
Hyperlinks = edges

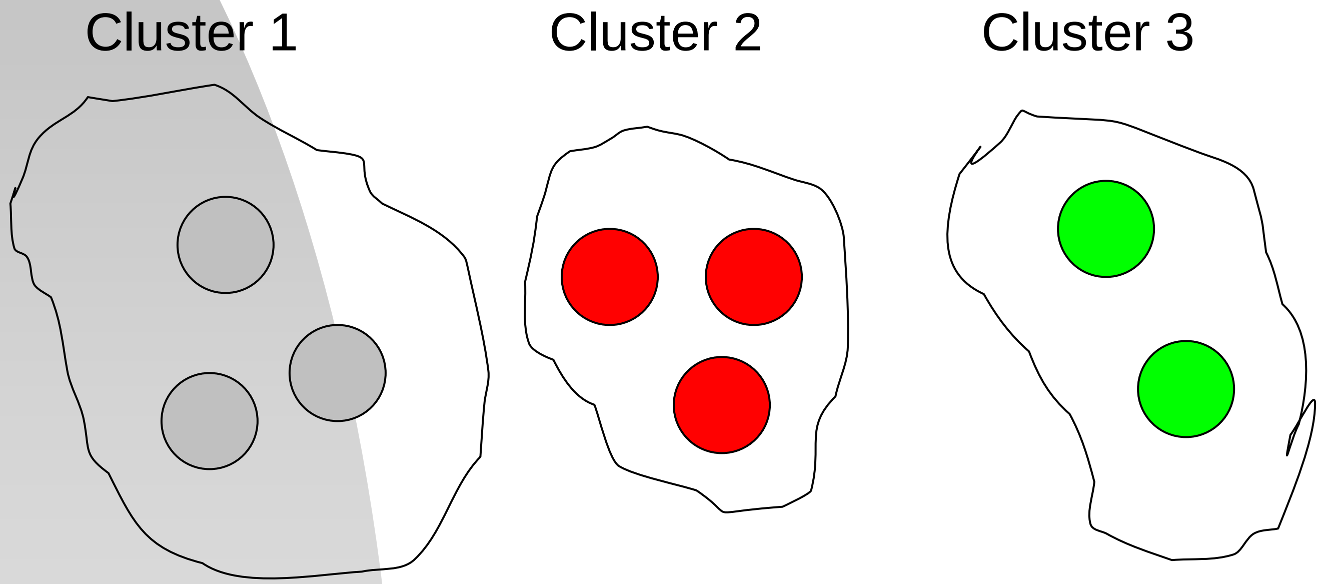
→

Huge and unreadable graphs!



Clustering

Clustering is a general technique aimed at gathering the entities that compose a system into cohesive groups (clusters).



Entities are grouped together when they possess similar properties.

The problem

- Can clustering of the pages composing a Web application be used to support program understanding?
- Several clustering techniques are available:

  - the pages can be described in different ways.

  - different similarity/distance measures are possible.

  - alternative algorithms can be used to form the clusters.

- The problem is how to evaluate the competing clustering techniques, in order to select the best (**if any**) for program understanding purposes.

Clustering techniques identified

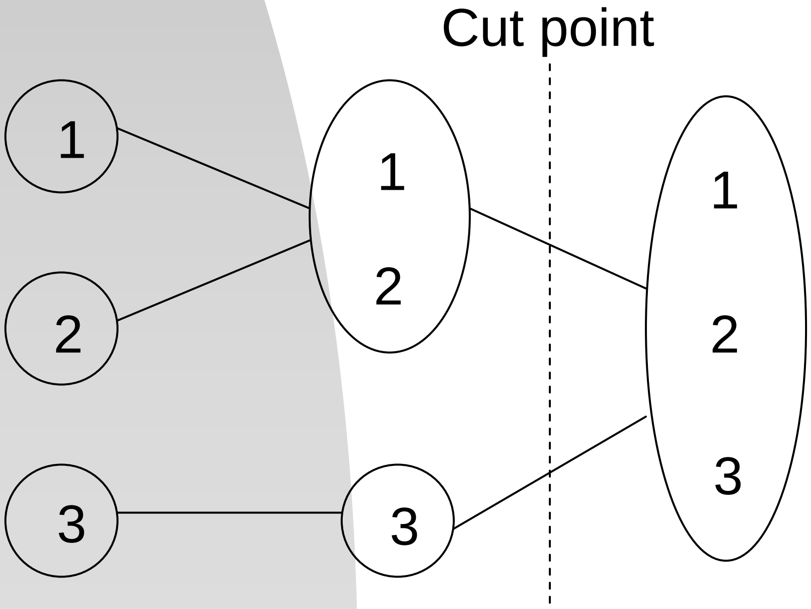
We have identified three alternative approaches that can be used to cluster Web pages.

page description - similarity/distance measure - algorithm

Structural:		
AST of the page	tree edit distance	agglomerative
Connectivity:		
Hyperlinks	“portions highly connected”	agglomerative
Keywords:		
Keywords	“common keywords”	agglomerative

Agglomerative hierarchical clustering

- This algorithm does not produce a single partition of the system but it builds a **hierarchy of clusterings**. Each level in the hierarchy defines a partition of clusters.
- To select the resulting clustering, a *cut point* has to be determined.



Hierarchy of clusterings for three entities.

The cut point determines 2 clusters, C1 = (1, 2) and C2 = (3).

Evaluation Methods

- Given the clusters produced by these three alternative approaches, the problem now is how to evaluate them.
- The result of clustering is a higher level view of a system. Such view may give useful information about the system or may be completely useless.
- There is no unique way to partition a system in a useful way, so that different clusterings of a Web application may be equally good and useful.
- We consider two complementary methods that can be used to evaluate the output of different clustering: **Gold standard** and **Task oriented approach**.

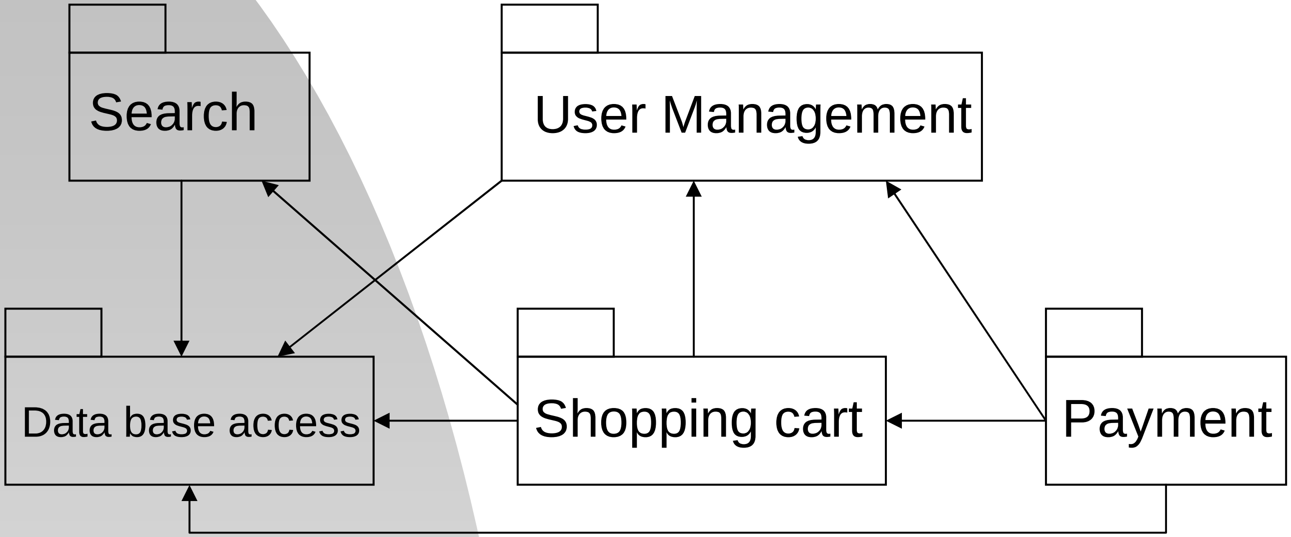
Gold standard

- The gold standard approach is a general evaluation method used to measure the performance of competing algorithms which compute a solution to a problem.
- The **gold standard** is the “ideal solution” to the problem.
- Usually the gold standard is determined manually by an expert on a set of examples.
- The competing algorithms are applied to such examples.
- The best algorithm is that which gives the solution closest to the “ideal” one.

Package diagram

- In UML the basic grouping mechanism that allows describing a system at a high level is called **package**.
- The **package diagram** is the related view.
- The package diagram gives the main components into which a system is logically divided.
- A package is a grouping of model elements (i.e., Web pages).
- Since clustering produces a grouping of Web pages, it makes sense to compare its output with the package diagram of the Web application under analysis

Package diagram of an e-commerce application



Packages contain groups of related pages:

Search: search.html, general-search.php, search-help.html, …
Data base access: query.php, db-lib.php, …
User management: registration.php, login.php, logout.php, …
Shopping cart: add-to-cart.php, del-from-cart.php, show.php, …
Payment: order.php, validate-credit-card.php, …

Gold standard is not sufficient!

- The package diagram is not the unique possible decomposition that can be used for Web application understanding.
- Alternative decompositions focused on specific aspects might be equally relevant.
- For this reason the gold standard approach need be complemented by a second evaluation method: **the task oriented approach**.

Task oriented approach

- The task oriented approach does not require that a correct output of the clustering technique be defined.
- If the output of a clustering method is helpful in conducting some activities in program understanding then the view extracted is considered meaningful.
- Some views may be useful for a category of tasks, while their support to tasks in other categories might be null.

Task oriented: expensive but fundamental complement

- Task oriented evaluations are **expensive**, because they require human intensive work in the definition and execution of the tasks, and in the scoring (assessment of the support provided).
- A task oriented evaluation is a **fundamental complement** to the gold standard:

  - it might be the case that the package diagram is not produced but the views recovered are a good support for program understanding.

  - it allows determining which clustering technique is more suited for which task (not provided by the gold standard).

Evaluation procedure: gold standard

1. Construction of the package diagram (if not available).
2. Computation of clustering by means of alternative techniques.
3. Clustering evaluation.

Clustering evaluation

In literature there exist different methods for comparing clusters with the gold standard. One of them is **precision/recall**.

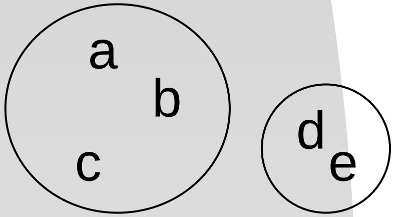
$$\text{Prec.} = \frac{\text{Num. of intra pairs in the test that are also in the gold}}{\text{total num. of intra pairs in test}}$$

$$\text{Rec} = \frac{\text{Num. of intra pairs in the gold that are also in the test}}{\text{total num. of intra pairs in gold}}$$

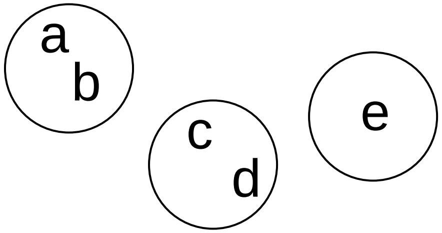
Example:

Precision = 1/4

Recall = 1/2



Test clustering



Gold clustering

Evaluation procedure: task oriented approach

1. Task definition.
2. Computation of clusters by means of alternative techniques.
3. Task execution.
4. Clustering evaluation.

Task definition

- The tasks used for clustering evaluation should be those typical of the activities performed by a Web developer during the evolution of a Web application.
- The best method for task definition would be interviewing the developer of the considered Web application and collecting a list of real tasks.
- If this is not possible, tasks should be determined “by playing” the role of the Web developer.

Task execution and clustering evaluation

Given a task, N+1 groups of Web programmers are necessary for evaluating N clustering methods .

Group 1 [Easy]

Output of clustering 1

Group 2 [Average]

Output of clustering 2

…

Group N+1 [Hard]

Without any support

To measure the support of each clustering techniques two Possible metrics are:

- time necessary to complete the task.
- subjective assessment on an ordinal scale of the level of difficulty encountered during the execution of the task.

Example: Tasks

1. Introduce a security check for all pages related to buying.
2. Remove the list of hyperlinks at the bottom of pages and replace them with a menu in a new frame.
3. Add links to similar products in each page describing a product.
4. Advertise the service of a given bank in each page related to the payment.

Conclusions (1)

- Two alternative approaches for the evaluation of the results produced by Web application clustering have been compared.
- Gold standard approach is appealing because it can be fully automated but it is not applicable if clustering is unable to reproduce a reference package diagram.
- The task oriented approach is expensive but has several remarkable advantages over the gold standard:

  - it allows determining which clustering technique is more suited for which task.

  - it gives information on the actual usefulness of each clustering technique.

Conclusions (2)

- The implementation of both approaches for the evaluation of a set of clustering techniques is essential to answer the question: “***can clustering support Web understanding and modification****?*”
- The ability of a clustering technique to recover the package diagram of a Web application is a strong indicator of a positive answer.
- In case of negative answer, the outcome of a task oriented empirical study could still indicate that the clustering views are useful, although not close to the package diagram.

Future Work

The implementation of:

- Gold standard approach

- Task oriented approach

for the evaluation of the following clustering techniques:

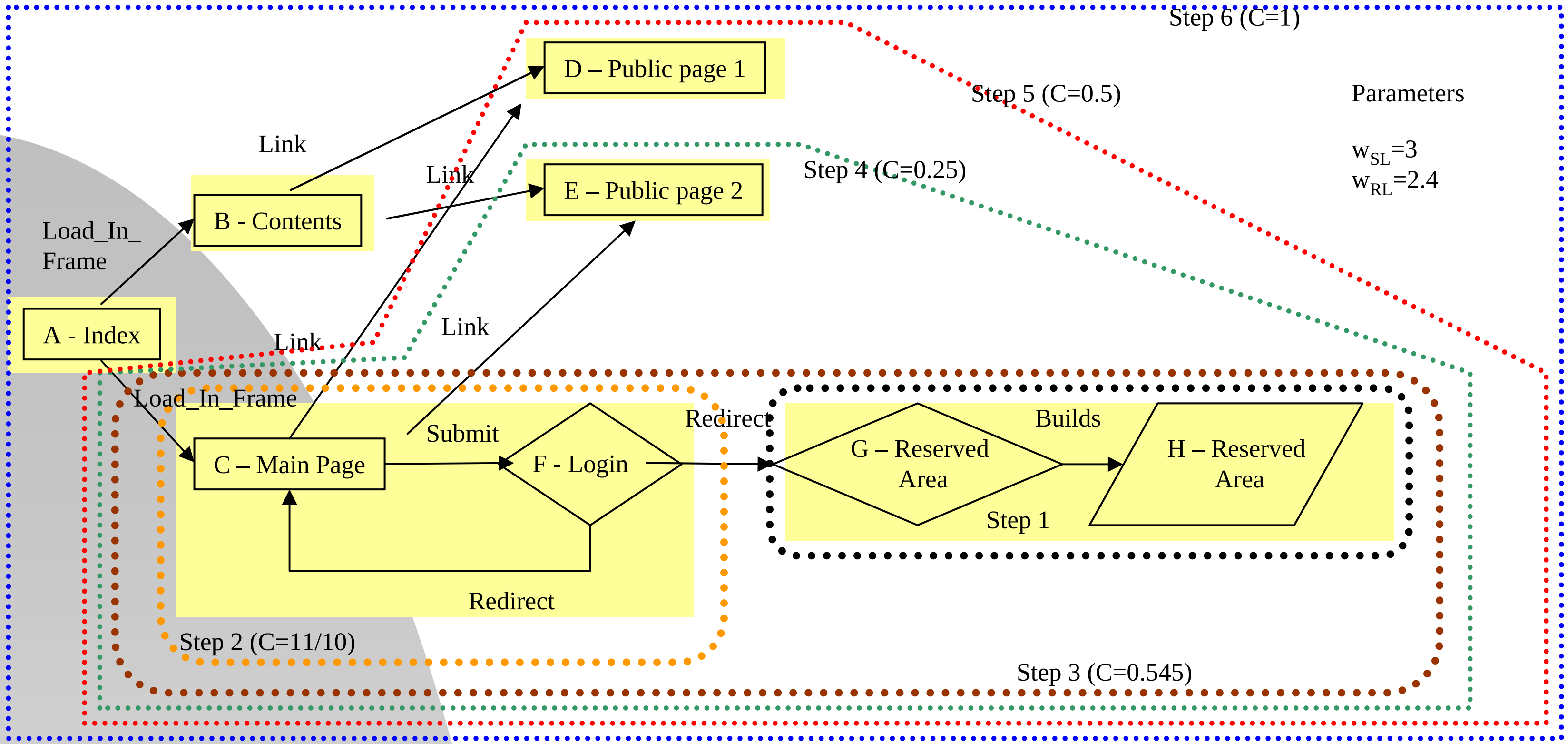
- Structural

- Connectivity

- Keywords

Connectivity Clustering tecnique

- Groups software components of a WA into meaningful (highly cohesive) and independent (loosely coupled) clusters.
- Evaluates the degree of coupling between interconnected components depending on both the typology and the topology of the connections,
- Proposes a clustering configuration that includes clusters with high intra-connectivity and low inter-connectivity
- Produces satisfying results in detecting components that collaborate for implementing a given functionality in a WA



Parameters

$w_{SL}=3$
$w_{RL}=2.4$

Step 2

$p_{B\to D}= p_{B\to E} =1/2$
$p_{C\to D}= p_{C\to E}= 1/5$
$p_{C\to F}=3/5$
$p_{F\to GH}=1$
$p_{F\to C}=1/2$

$p_{C\leftarrow F}=1$
$p_{F\leftarrow C}=1$
$p_{GH\leftarrow F}=1$
$p_{D\leftarrow B}= p_{E\leftarrow B}= 1/2$
$p_{D\leftarrow C}= p_{E\leftarrow C}= 1/2$

$C_{B,D}= C_{B,E} =1/2*1/2=1/4$
$C_{C,D}= C_{C,E}=1/5*1/2=1/10$
$\mathbf{C_{C,F}=3/5*1+1/2*1=11/10}$
$C_{F,GH}=1*1=1$
QoC=0.132143

Step 3

$p_{B\to D}= p_{B\to E} =1/2$
$p_{CF\to D}= p_{CF\to E}= 1/4.4$
$p_{CF\to GH}=0.545$

$p_{GH\leftarrow CF}=1$
$p_{D\leftarrow B}= p_{E\leftarrow B}= 1/2$
$p_{D\leftarrow CF}= p_{E\leftarrow CF}= 1/2$

$C_{B,D}= C_{B,E} =1/2*1/2=1/4$
$C_{CF,D}= C_{CF,E}=1/4*1/2=1/8$
$\mathbf{C_{CF,GH}=0.545*1=0.545}$
QoC=-0.023889

Step	*QoC*	*C*
1	-0,11607	
2	0,132143	1,1
3	-0,02389	0,545
4	0.127083	0,25
5	0.022778	0,5
6	0,055556	1

The best QoC is for 0.545<Cut height<1.1