

# Tecniche di personalizzazione di interfacce Web

Parte dei contenuti delle seguenti slides sono stati estratti da “Tasso C., Omero P, La personalizzazione dei contenuti WEB, Franco Angeli, 2002” ed arricchite da altre fonti per finalità puramente didattiche.

# Tre fasi principali per la personalizzazione

- riconoscere l'utente che accede al sito - associarlo con un profilo contenuto in un archivio;
- identificare i bisogni e le caratteristiche dell'utente;
- fornire all'utente i contenuti ritenuti per l'utente stesso più rilevanti

The screenshot displays a web browser window with the address bar showing <http://www.webgroove.com/demo/>. The page content includes a navigation sidebar on the left with the 'relate' logo and a 'Demo' section with steps: 1. Read This Introduction, 2. Click Staff News Example Share Example, 3. Peek Inside Analysis, 4. Experiment!, and a 'See Also' section with links like 'Outlier Relationships' and 'Whitepapers'. The main content area is titled 'Other News' with the subtitle 'For the Stories You May Have Missed' and a 'Top Stories' section. This section is divided into four columns: Politics (Clinton turns helm over to shock jock, Gen Xers cheer, Hollywood celebs vying for cabinet positions, Women vow to form separate nation if Stem steps), Business (Stocks plummet as Howard Stern wins White House, IBM loses stronghold after Y2K crash, Last two banks merge, citizens fear), Entertainment (Denise Rodman to wed Madonna, baby 3 on the way, Simpson in 3rd season, still fresh?, New Cameron flick sets record, \$1 billion in first weekend), and Sports (Sosa and McGwire come to blows in Chicago restaurant, Cowboys top boy arrested in sex scandal, Michael Jordan retires again). At the bottom, there is a 'relate' logo, a bar chart, and three buttons: ANALYSIS (green), DECAY (orange), and CLEAR (red).

## Metodi di identificazione dell'utente: il cookie

**Il cookie:** file di piccole dimensioni trasmessi dal WEB Server del sito internet e memorizzati nel computer utente

### **Procedura:**

- 1- l'utente X accede per la prima volta ad un sito Y;
- 2- il sito scarica sul computer dell'utente X un cookie;
- 3- l'utente si ricollega con lo stesso computer al sito Y, il sistema identifica la macchina dell'utente in virtù del cookie precedentemente scaricato

- a) il sistema identifica una macchina non un utente (se l'utente si collega da un altro computer non viene riconosciuto);
- b) il metodo non è sempre applicabile (cancellazione del cookie ad opera dell'utente e/o modifica delle impostazioni del browser)

# (I) Metodi di identificazione dell'utente: user-id e password

Richiesta di user-id e password: il sistema identifica lo specifico utente e non la sua macchina (l'utente può collegarsi da un computer diverso).

## **Applicabilità del metodo:**

- si può applicare solo se l'utente è disponibile a registrarsi;
- alcuni utenti rifiutano di fornire i dati previsti dalla compilazione;
- l'utente ha paura di ricevere nuove informazioni e rinuncia al servizio;
- possono esistere utenti che forniscono informazioni false

## (II) Metodi di identificazione dell'utente: Agenti software

Sono software che gestiscono accessi ai servizi di personalizzazione (e altro) - NewsAlert è uno esempio

Viene riconosciuto il calcolatore  
(stessi problemi del cookie)

Applicabilità del metodo:

- sono legate alle difficoltà da parte di un utente ad installare software

# Identificazione dei bisogni e delle caratteristiche individuali dell'utente

Acquisizione di informazione sull'utente:

- **esplicita;**
- **implicita**

Sull'esplicita nessun problema

## Metodi di acquisizione implicita, Web Usage Mining: il tracking dell'utente

Il tracking permette di osservare ed analizzare il comportamento dell'utente attraverso dati prodotti dal Web Sever riguardanti gli accessi operati dall'utente:

- tutte le azioni compiute all'utente;
- ricostruzione dell'intera sessione;
- intera sequenza di pagine richieste dall'utente nello stesso sito (intervallo non superiore ad un numero fissato di minuti - 30 circa )

## Ambiti applicativi per il Web Usage Mining relativi alla personalizzazione

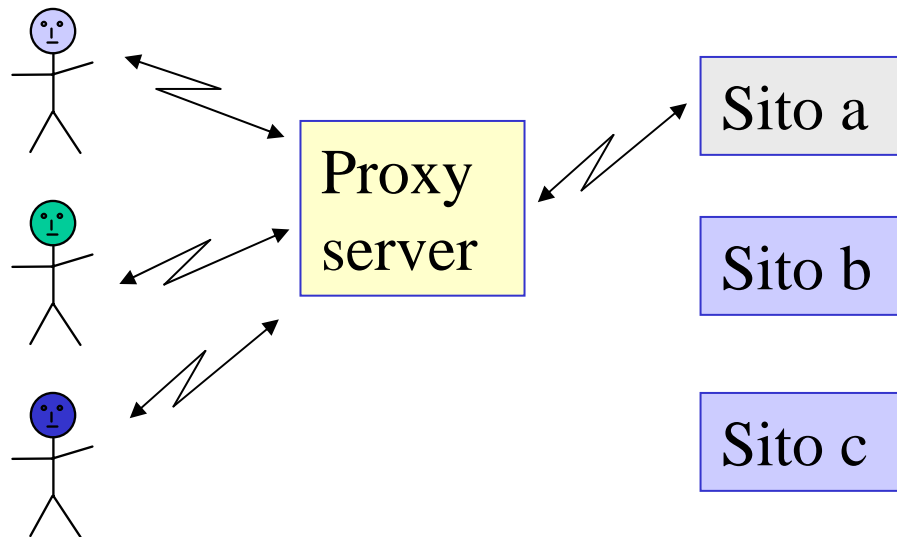
- Acquisire informazioni con cui raffinare ed estendere il profilo utente;
- Scoprire tendenze di acquisto dell'utente;
- Scoprire correlazioni fra contenuti;
- Utilizzare tecniche di personalizzazione effimera;
- Personalizzare il sito

## Primo problema - il local caching

**Local caching:** i programmi per visualizzare pagine WEB hanno una memoria a disposizione (la cache) - per motivi legati all'efficienza - se una pagina richiesta è presente nella cache non viene inoltrata al sito – di conseguenza potrebbero mancare delle richieste fatte (pagine) nella costruzione del percorso all'utente

## Secondo problema: il proxy server

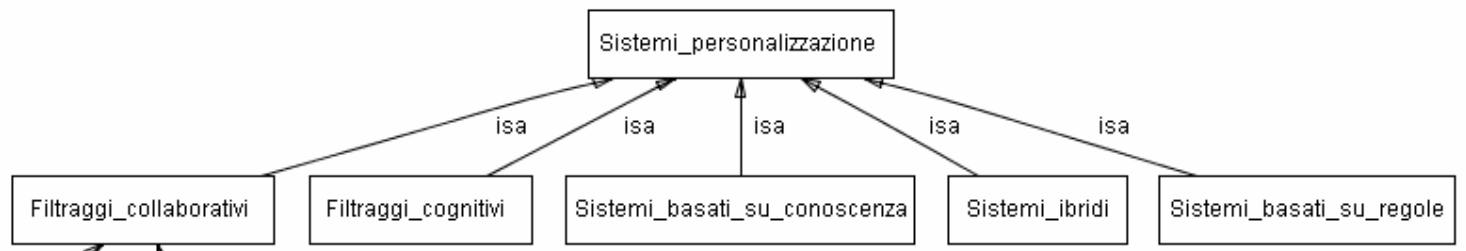
Il proxy è un computer che si interpone fra utenti e il WEB server - sul proxy vengono memorizzate le pagine WEB che gli utenti hanno visitato - quando un utente richiede una pagina WEB se è presente in quelle memorizzate, il proxy la fornisce



Problemi: a) le macchine dell'utente che si avvalgono di un proxy appaiono identificati da un unico IP (l'identificazione fallisce); b) si verifica lo stesso problema del browser della propria cache locale

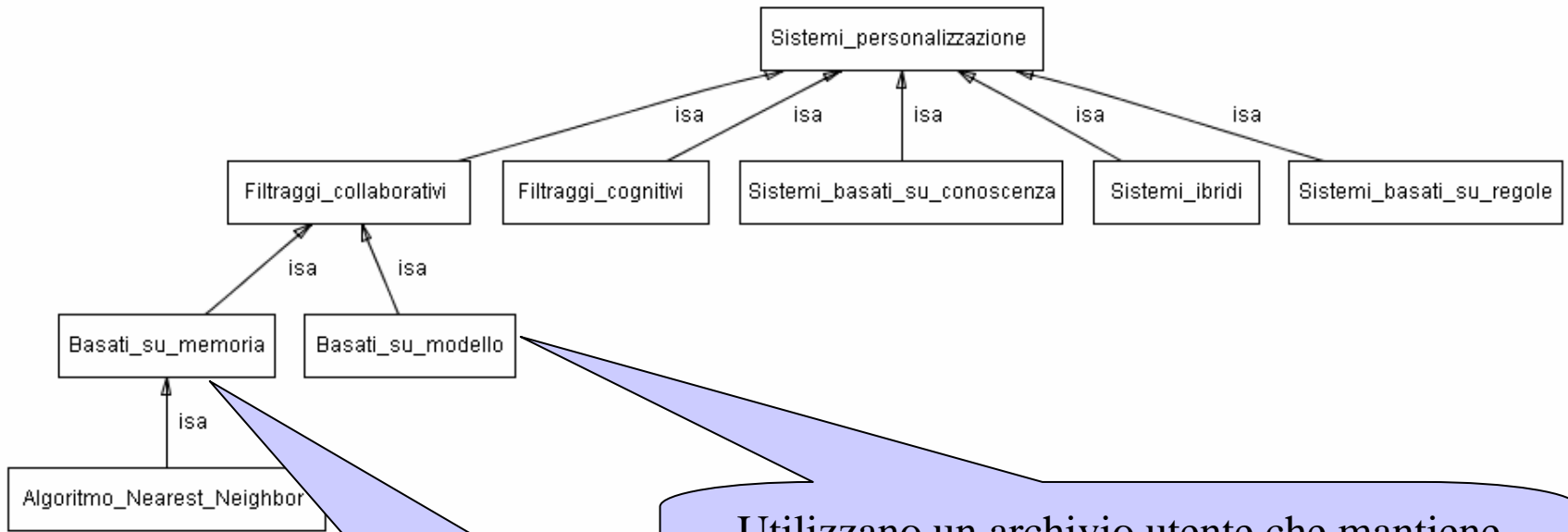
## Il processo di personalizzazione non viola la privacy

- Nessun metodo identificazione e tracking fornisce informazione sulla reale identità della persona che si collega ad un sito;
- i sistemi che utilizzano dati personali che identificano l'utente debbono attenersi a vicoli normativi.



# Una prima classificazione dei sistemi di personalizzazione

# Filtraggio su memoria e su modello



Utilizzano un archivio utente che mantiene le correlazioni fra gli utenti

Utilizzano un archivio utente contenenti i voti di recensione di ciascun utente e calcolano previsione su queste basi

# Filtraggio collaborativo

utenti	contenuti							
	C1	C2	C3	C4	C5	C6	C7	C8
U1	4		9	3	3			
U2	6	1	8		2			
U3								
U4		2		3				
U5			3		5			
U6		7				6		

Basato su una matrice di recensioni in cui vengono riportati indicazioni (valori qualitativi - quantitativi) sui gusti degli utenti - l'acquisizione dei valori di recensione può essere esplicita o implicita.

Principio di base: se due utenti hanno mostrato gli stessi interesse nel passato probabilmente continueranno a mantenerli anche nel futuro

# Algoritmo Nearest Neighbor

Sarwar et. Al. 2000 <http://citeseer.nj.nec.com/345989.html>

		contenuti							
		C1	C2	C3	C4	C5	C6	C7	C8
utenti	U1	4		9	3	3			
	U2	6	1	8		2			
	U3								
	U4		2		3				
	U5			3		5			
	U6		7					6	

1. Calcolo della similitudine fra utenti;
2. Identificazione delle guide;
3. Calcolo della previsione;
4. Normalizzazione del voto di previsione

# Calcolo della similitudine fra utenti

Ad esempio lo scarto quadratico medio

$$sim(u_a, u_k) = \frac{\sum_{i=1}^n (v_{ki} - v_{ai})^2}{n}$$

L'esempio della matrice di recensioni precedente

$$sim(u_1, u_2) = \frac{(4 - 6)^2 + (9 - 8)^2 + (3 - 2)^2}{3} = 2$$

# I dentificazione delle guide

- a) Metodo della soglia
- b) Metodo degli  $n$  vicini migliori

## 3. Calcolo delle previsioni

Si considerano le recensioni delle guide per formulare la previsione: più una guida è simile ad un utente più la recensione utilizzata è ritenuta importante per quell'utente

# Normalizzazione

Nasce dal problema della diversa percezione degli utenti e dalla diversa scala di valori utilizzati per le recensioni

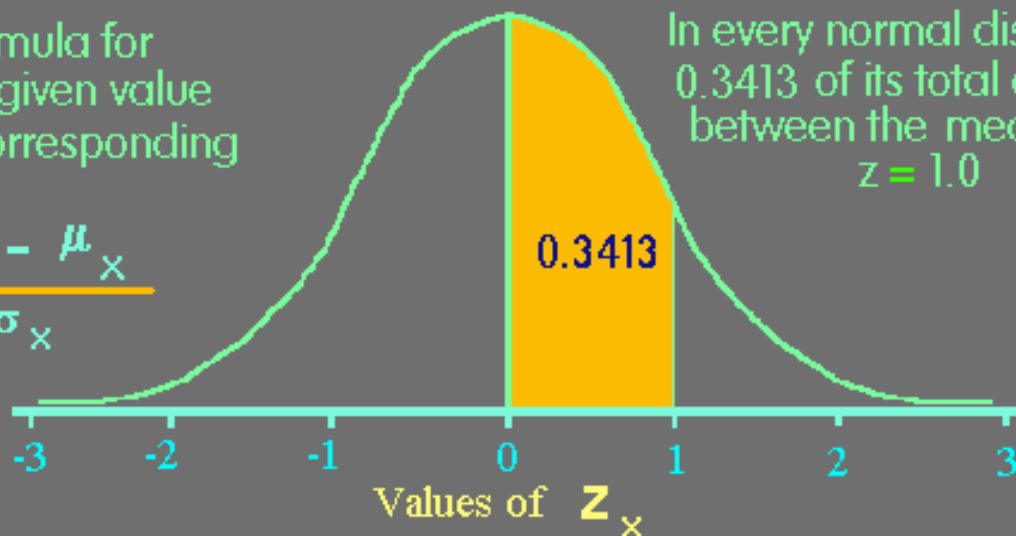
Z scores are a special application of the transformation rules. The z score for an item, indicates how far and in what direction, that item deviates from its distribution's mean, expressed in units of its distribution's standard deviation. The mathematics of the z score transformation are such that if every item in a distribution is converted to its z score, the transformed scores will necessarily have a mean of zero and a standard deviation of one.

Z scores are sometimes called "standard scores". The z score transformation is especially useful when seeking to compare the relative standings of items from distributions with different means and/or different standard deviations.

Z scores are especially informative when the distribution to which they refer, is normal. In every normal distribution, the distance between the mean and a given Z score cuts off a fixed proportion of the total area under the curve. Statisticians have provided us with tables indicating the value of these proportions for each possible Z score.

This is the formula for converting a given value of  $x$  into its corresponding z score:

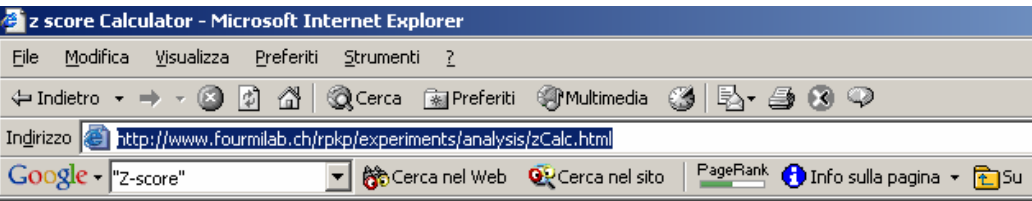
$$Z_x = \frac{X - \mu_x}{\sigma_x}$$



In every normal distribution 0.3413 of its total area lies between the mean and  $z = 1.0$

All'indirizzo:

<http://www.fourmilab.ch/rpkp/experiments/analysis/zCalc.html>



## z score Calculator

The probability of a result  $x$  in an experiment consisting of a large number of equally probable independent trials  $n$  is

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{2\sigma^2}\right)^2}$$

where  $\mu$ , the *mean value*, is  $n/2$  and  $\sigma$ , the *standard deviation*, is a measure of the breadth of the curve which, for ex

$$\sigma = \sqrt{\mu/2} = \sqrt{n/4}$$

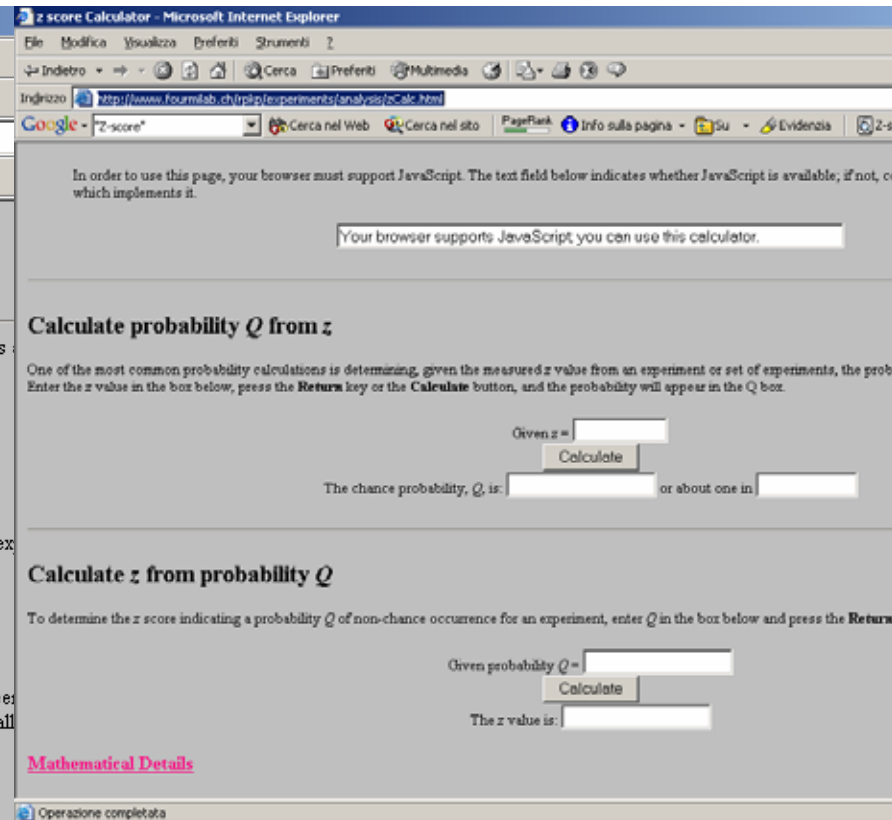
Since the mean value and standard deviation depend upon the number of trials in the experiment, comparison between *standardising* the result: transforming it to a distribution with mean value zero and standard deviation of 1. A normal subtracting the mean and dividing by the standard deviation of the experiment:

$$z = \frac{x - \mu}{\sigma}$$

This *z-value* or *z score* expresses the divergence of the experimental result  $x$  from the most probable result  $\mu$  as a number of standard deviations  $\sigma$ . The larger the value of  $z$ , the less probable the experimental result is due to chance. The probability can be calculated from the *cumulative standard normal distribution*:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

Which gives the probability  $P$  that an experimental result with a  $z$  value less than or equal to that observed is due to chance. Subtracting  $P$  from one:



In order to use this page, your browser must support JavaScript. The text field below indicates whether JavaScript is available, if not, which implements it.

Your browser supports JavaScript you can use this calculator.

### Calculate probability $Q$ from $z$

One of the most common probability calculations is determining, given the measured  $z$  value from an experiment or set of experiments, the probability  $Q$ . Enter the  $z$  value in the box below, press the **Return** key or the **Calculate** button, and the probability will appear in the  $Q$  box.

Given  $z =$

The chance probability,  $Q$ , is:  or about one in

### Calculate $z$ from probability $Q$

To determine the  $z$  score indicating a probability  $Q$  of non-chance occurrence for an experiment, enter  $Q$  in the box below and press the **Return** key or the **Calculate** button.

Given probability  $Q =$

The  $z$  value is:

### Mathematical Details

Operazione completata

Operazione completata

Internet

# L'algoritmo di LikeMinds in MovieCritic

**MovieCritic.ca**

**AllPosters.com** [Click Here](#)

[HOME](#) | [REVIEWS & RATINGS](#) | [BEST & WORST](#) | [NEWS](#) | [FAQ](#) | [CONTACT](#) | [LINKS](#) | [FUN STUFF](#) | [POSTERS](#)

"Survey says...." [Click here!](#) And put in your two cents worth!

## Featured Movie Reviews



[Windtalkers](#)



[Spider-Man](#)



[Deuces Wild](#)



[Panic Room](#)

### Search

[Advanced Search](#)

### Mailing List

Receive my reviews by e-mail!

[Privacy Policy](#)

## Latest Movies I Have Rated

Title	Rating	#
<a href="#">Windtalkers (2002)</a>	C-	1515
Spirit: Stallion of the Cimarron (2002)	B	1514
The Sum of All Fears (2002)	B	1513
Star Wars: Episode II - Attack of the Clones (2002)	A	1512
About a Boy (2002)	B	1511
Life or something like it (2002)	C+	1510

## Upcoming Movie Releases

September 6, 2002	
City by the Sea	<a href="#">[IMDb]</a>
Swimfan	<a href="#">[IMDb]</a>
September 13, 2002	
Barbershop <a href="#">[Poster]</a>	<a href="#">[IMDb]</a>
Stealing Harvard	<a href="#">[IMDb]</a>
The Transporter	<a href="#">[IMDb]</a>
September 20, 2002	

## Featured Products

[Check out my College Posters Store!](#)

## What's new?



# Le fasi dell'algoritmo Nearest Neighbor: LikeMinds in Moviecritic (I)

1. Fase di identificazione;
2. Fase di recensione dei film e costruzione dei voti;
3. Fase di identificazione di un insieme di utenti che abbiano valutato alcuni film comuni;
4. Fase di calcolo della similitudine fra utenti (valore di somiglianza);

$$AS(u_k, u_m) = \frac{\sum_{i=1}^n p(v_{ki} - v_{mi})}{n} \log_2 n$$

# Le fasi dell'algoritmo Nearest Neighbor: LikeMinds in Moviecritic (II)

5 Fase di identificazione della guida (utente più simile)

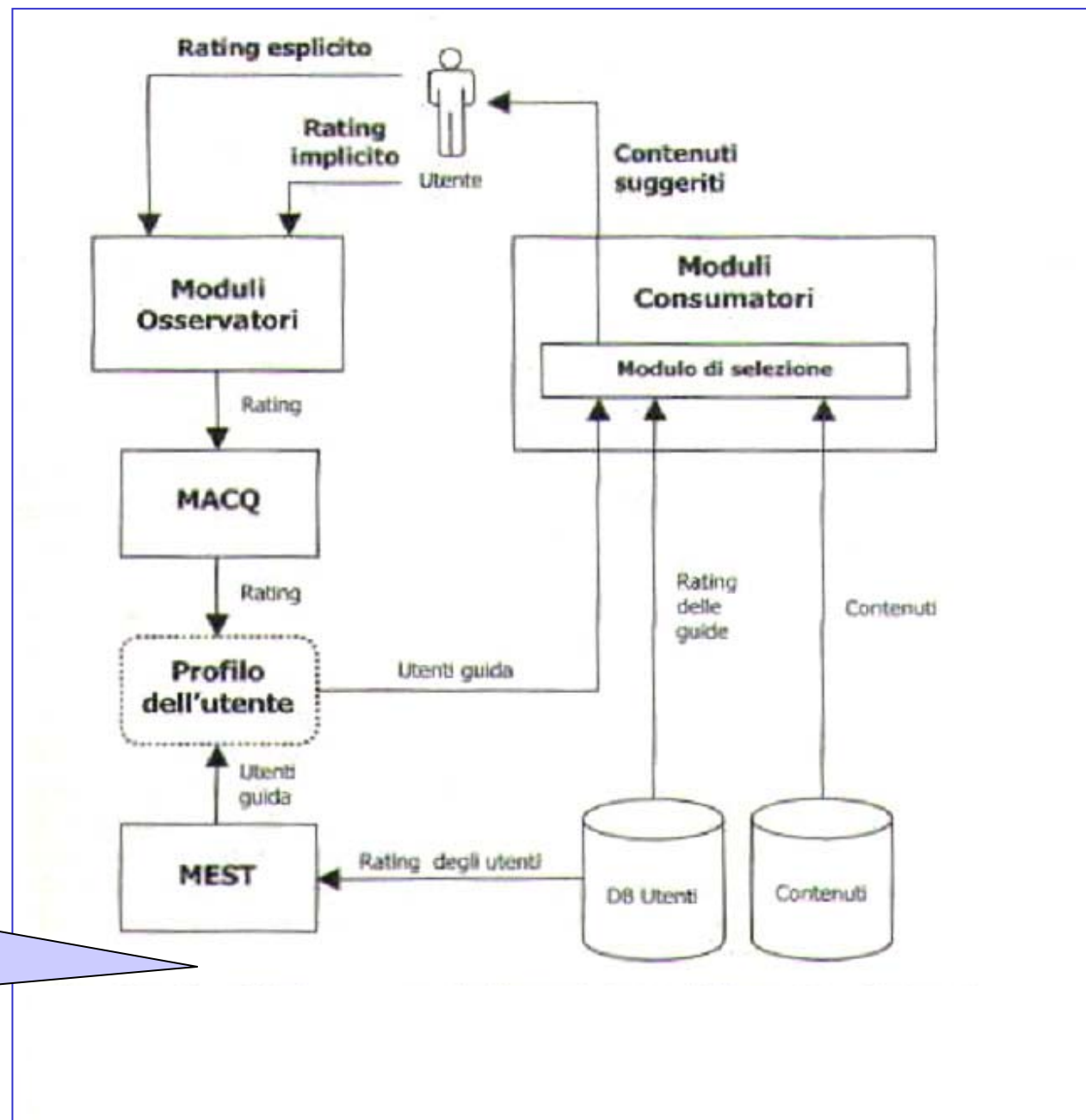
6 Fase del calcolo della raccomandazione o della previsione (due versioni)

a) previsioni derivate dalle guide con peso maggiore;

b) previsione calcolata con qualche funzione applicata alle guide;

7 Fase di normalizzazione (previsione modificata in accordo con la scala dei valori dell'utente attivo)

# Architettura filtraggio collaborativo



Modulo di estensione  
dal db utente + similitudine +  
previsioni

# Vantaggi del filtraggio collaborativo

- Indipendenza della natura dei contenuti trattati (si basano unicamente sulle opinioni degli utenti);
- Tutela della privacy e dell'anonimato degli utenti (è necessario solo il login e password);
- Incremento progressivo delle prestazioni;
- Possibile identificazione di nuove aree di interesse per l'utente

# Svantaggi del filtraggio collaborativo

- La partenza a *freddo*
- Il problema della dispersione delle recensioni (archivio molto esteso)
- La soggettività di giudizio (incapacità di recensione omogenea);
- Mancanza di motivazione a compilare le recensioni;

http://www.villagenet.com/Directory/Computers/Software/Internet/Servers/Personalization/

Personalization - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Indirizzo <http://www.villagenet.com/Directory/Computers/Software/Internet/Servers/Personalization/>

Google personalization download Cerca nel Web Cerca nel sito PageRank Info sulla pagina Su Evidenzia personalization download

home	forums	chat	e-mail	shop	help	sign-up
search the internet	local news	world news	local business news	sports		

search the internet

stocks: enter ticker


television: enter ZIP

movies: enter ZIP

**villageworld.com services**

citizenship

welcome to **villageworld.com** of new york



## Personalization

[Root](#) / [Computers](#) / [Software](#) / [Internet](#) / [Servers](#) / Personalization

- [Aptex Software](#)  
Using proprietary Content Mining technology, Aptex develops and markets intelligent solutions for online publishing, customer response, and educational publishing.
- [BroadVision](#)  
BroadVision is a supplier of Internet application solutions for one-to-one relationship management across the extended enterprise.
- [DataSage](#)  
Netcustomer application analyzes large amounts of customer data gathered from online and offline marketing channels to enable individualized Internet marketing. Tracks customer behavior to individualize site content, banner ads, email and other direct communications.
- [WebFalcon](#)  
Interactively targets information to web site visitors based on a visitor profile which is determined by their selections while navigating your web site.
- [Open Profiling Standard \(OPS\)](#)

**News example page: click on a few links below to affect your profile**



# Other News

For the Stories You May Have Missed

## Top Stories

### Politics

- [Clinton turns helm over to shock-jock, Gen X'ers cheer](#)
- [Hollywood celebs vying for cabinet positions](#)
- [Women vow to form separate nation if Stern stays](#)

### Business

- [Stocks plummet as Howard Stern wins White House](#)
- [IRS loses stranglehold after Y2K crash](#)
- [Last two banks merge, citizens fear](#)



### Entertainment


- [Dennis Rodman to wed Madonna, baby 3 on the way](#)
- [Simpsons in 32nd season, still fresh?](#)
- [New Cameron flick sets record: \\$1 billion in first weekend](#)

### Sports

- [Sosa and McGwire come to blows in Chicago restaurant](#)
- [Cowboys top boy arrested in sex scandal](#)
- [Michael Jordan retires, again](#)

- Demo**
- 1. Read This Introduction
  - 2. Click Stuff News Example Store Example
  - 3. Peek Inside Analysis
  - 4. Experiment!
- See Also**
- Overview
  - Relationships
  - Whitepapers
  - Home
  - About Us
  - Make Contact





**Demo**

- [1. Read This Introduction](#)
- [2. Click Stuff News Example Store Example](#)
- [3. Peek Inside Analysis](#)
- [4. Experiment!](#)

**See Also**


- [Overview](#)
- [Relationships](#)
- [Whitepapers](#)
- [Home](#)
- [About Us](#)
- [Make Contact](#)

### What We Learned About You

<b>Mildly Interested</b>	music, travel, football, movies
<b>Interested</b>	basketball, science, computers, health, sports, entertainment, politics
<b>Very Interested</b>	finance, business

### Top 10 Recommended Ads

Peak Analysis	Threshold Analysis
<b>All Business News, All the Time</b>	<b>All Business News, All the Time</b>
Stay on top of your savings	Stay on top of your savings
<b>Internet Sports News</b>	We've got the inside scoop
We've got the inside scoop	<i>Get the inside scoop on your favorite stars</i>
<b>We ARE football on the Web</b>	<b>Internet Sports News</b>
<i>Get the inside scoop on your favorite stars</i>	 We've got what you want



WGR PROFILER PO-S PROFILE DISPLAY

TEST INPUTS: ANALYSIS (green), DECAY (yellow), CLEAR (red)



## Demo

[1. Read This Introduction](#)[2. Click Stuff](#)[News Example](#)[Store Example](#)[3. Peek Inside Analysis](#)[4. Experiment!](#)

## See Also

[Overview](#)[Relationships](#)[Whitepapers](#)[Home](#)[About Us](#)[Make Contact](#)

your greatest interests for matching purposes.

## Top 10 Recommended Books

### Peak Analysis

Maximizing Your Stock Portfolio  
 Guide to Shady Dealmaking  
 Inside Microsquash Office Professional  
 Travel Expense Logbook  
 Making it as a Soap Opera Actor  
 How to Become President in 21 Days  
 Race to the White House  
 Through the Wizard's Gate  
 Musician's Market 2000  
 Technophobia

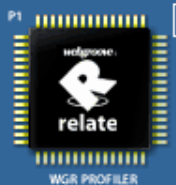
Peak analysis makes a fuzzy comparison between individual peak interests and product profiles. Customers get the most relevant recommendations based on available products. This is great for most online commerce situations.

### Threshold Analysis

No matching items found.

Threshold analysis can be used for highly targeted product placement. A marketer can specify exact interest score thresholds, so his products are only shown to people who meet that criteria, increasing point-of-purchase sales.

**Next Step:** go back and [click more](#) or clear your profile and start over



WGR PROFILER



PO-S PROFILE DISPLAY



ANALYSIS

DECAY

CLEAR