

Sistemi per il recupero delle informazioni

Gabriele Pozzani

A.A. 2013/2014

**Corso di Laurea Magistrale in
Editoria e Giornalismo**

La ricerca personalizzata

basato sulla tesi

**“Google e la ricerca personalizzata: opportunità e limiti tra
consapevolezza e ignoranza” della Dott.ssa Rachele Giorgetti**

Cosa è

- E' il processo attraverso il quale un motore di ricerca ritorna risultati rilevanti per **l'utente** che ha effettuato una specifica interrogazione
 - Quello specifico utente

3

Personalizzata e Generica (I)

Ricerca generica (non personalizzata)

- È “oggettiva” perché appiana le differenze fra individui: “one size fits all”;
- Non costruisce profili utente ma ritorna a tutti gli utenti che effettuano la stessa query gli stessi risultati
- Utilizza criteri “oggettivi” per ordinare i risultati (Ranking). Es: PageRank

Ricerca personalizzata

- È “soggettiva” perché esalta le differenze fra individui: “one size for one person”;
- Costruisce profili utente per ritornare documenti su misura ad ogni utente.
- Utilizza criteri “soggettivi” per ordinare i risultati (Ranking). Es: Profili utente

4

Personalizzata e Generica (II)

- N.B.: Ricerca generica e ricerca personalizzata NON si escludono a vicenda ma si integrano reciprocamente
 - Prima di tutto, il ranking dei risultati avviene sempre secondo criteri oggettivi
 - In un secondo momento, interviene la ricerca personalizzata, la quale filtra i risultati da ritornare all'utente.

5

Perché la personalizzazione?

- Scopo dichiarato:
 - È il miglioramento delle performance del motore: la ricerca personalizzata intende facilitare il lavoro dell'utente, il quale riesce più velocemente a trovare le informazioni desiderate
- Scopo reale:
 - “Aiutare” le aziende a programmare le proprie inserzioni pubblicitarie sulla base dei singoli profili utente

6

Il profilo utente (I)

- Per poter personalizzare la ricerca sull'utente, l'utente va conosciuto:
 - Interessi
 - Conoscenza, background e abilità
 - Obiettivi
 - Comportamento
 - Caratteristiche individuali
 - Contesto in cui l'utente è inserito

7

Il profilo utente (II)

- È la descrizione dei più importanti o interessanti fatti/tratti relativi ad ogni singolo utente
- Le informazioni raccolte per costruire il profilo utente possono essere:
 - Fornite esplicitamente dall'utente, 3 problematiche:
 - Gli utenti non forniscono informazioni in modo esauriente
 - Spesso le informazioni fornite non corrispondono al vero
 - Incapacità di esprimere i propri interessi
 - Raccolta implicita
 - Raccolte attraverso l'osservazione delle sue azioni

8

La raccolta esplicita

- In Google, basata sulla creazione di un account
 - Nome e cognome
 - Data di nascita
 - Sesso
 - Cellulare
 - Indirizzo e-mail
 - Paese
- Questi dati personali sono richiesti per proteggere il proprio account e per assicurare all'utente un'esperienza “ottima”, indipendentemente dal servizio Google utilizzato
<http://support.google.com/accounts/bin/answer.py?hl=it&answer=173322>
- N.B. La raccolta esplicita include anche tutte le informazioni che l'utente inserisce durante l'utilizzo dei servizi Google

9

La raccolta implicita

- Tramite l'osservazione del comportamento di un utente
 - Argomenti delle interrogazioni
 - Formulazione e parole usate nelle interrogazioni
 - Click sui risultati
 - Servizi usati
 - ...
- lo si può “catalogare” attraverso metodi probabilistici
 - Bayesian Networks
 - Association Rules
 - Case-Based Reasoning
- N.B.: Questi sistemi forniscono risultati probabilistici, non certi, e si integrano fra loro ad ogni ricerca.

10

Cosa dice la homepage della “Guida di Google” sulla personalizzazione

“Tramite la personalizzazione dei risultati in base alla tua cronologia delle ricerche, ci auguriamo di poterti offrire i contenuti più utili e pertinenti per la tua ricerca. La personalizzazione della cronologia delle ricerche è soltanto uno dei metodi che ci consentono di visualizzare risultati di ricerca più personalizzati”

(<http://support.google.com/accounts/bin/answer.py?hl=it&answer=54041>)

11

Quando e come vengono raccolte le informazioni

CON ACCOUNT

- Ogni utente ha il proprio account
- Nell'account vengono raccolte **diverse informazioni** senza limiti di tempo

SENZA ACCOUNT

- Ogni browser ha dei “cookie” che permettono di “identificarlo”
- Google registra “solo” tutte le **ricerche** svolte negli ultimi 180 giorni
 - Query effettuate
 - Risultati cliccati
- Problema
 - Più persone potrebbero usare lo stesso browser sullo stesso PC
 - La personalizzazione è più limitata (nel tempo e nelle funzionalità)

12

Google dashboard

- Punto di accesso di un utente con account al proprio “dossier”
<https://www.google.com/dashboard>
- Le informazioni sono raggruppate per tipologia/servizio
 - Informazioni personali
 - Dispositivi (tablet, cellulare, ...)
 - Calendari
 - Contatti
 - Cronologia web (i.e., delle ricerche, tutte: pagine web, immagini, video, ...)
 - Cronologia delle posizioni (geografiche)
 - Email
 - Post su G+
 - Libri
 - Mappe
 - Foto
 - Acquisti su Play Store
 - Siti più visitati e bookmarks (se si usa Google Chrome)
 - Siti web
 - conversazioni su Google Talk
 - Video (YouTube)

13

Cronologia web

- Memorizza tutte le ricerche effettuate, permettendo di
 - Visualizzare e cercare tra pagine web visitate in passato
 - Individuare tendenze: esempio i siti più visitati e le ricerche più frequenti
 - Ricevere sui prodotti e servizi Google risultati di ricerca e consigli più personalizzati
 - Visualizzare suggerimenti personalizzati in completamento automatico
 - Filtrare i risultati di ricerca

14

I risultati personali

- La personalizzazione permette anche di includere nei risultati delle ricerche “risultati personali”
 - Contenuti personali dell’utente;
 - Contenuti condivisi da connessioni suggerite che hanno profili Google+
 - Connessioni suggerite:
 - Connessioni reciproche con persone in prodotti di Google
 - Interazioni avute con altre persone in altri prodotti di Google
 - Link aggiunti al proprio profilo
 - Account collegati associati al proprio account Google
 - Persone che fanno parte delle tue cerchie estese
 - Amici su siti esterni che hanno collegato i propri account ai profili Google+
 - Contenuti raccolti dall’utilizzo dei servizi.

15

Oltre alla personalizzazione, con questi dati

- Sostituisce nomi associati in precedenza all’account in modo da rappresentare in modo uniforme l’utente su tutti i servizi
- Utilizza il nome specificato dall’utente per il profilo Google in tutti i servizi offerti che richiedono un account
- Qualora altri utenti conoscessero già l’indirizzo e-mail di un utente o altre informazioni che lo identificano, mostra loro le informazioni del profilo visibili pubblicamente, come nome e foto
- Quando gli utenti contattano Google, registra la comunicazione per risolvere eventuali problemi
- Utilizza l’indirizzo e-mail dell’utente per fornirgli informazioni sui servizi, comunicargli i prossimi cambiamenti o miglioramenti

16

Ma...

- Google fornisce la possibilità di disabilitare la personalizzazione
 - Uscire da Google (log out);
 - Eliminare o sospendere definitivamente tutti (o singoli) gli elementi dalla Cronologia web (e altre “raccolte dati”) in qualsiasi momento
 - diversi sistemi permettono di navigare anonimi sulla rete

17

Navigare anonimi

- Non lasciare tracce di visita sui siti
- Non farsi localizzare
 - nascondere il luogo da cui ci si collega
- Sfruttare servizi online oscurati o bloccati
- Si usano servizi/sistemi chiamati “proxy”

18

Che cos'è un proxy?

- È un programma/server
- Si interpone tra un client e un altro server facendo da tramite o interfaccia tra i due
 - Inoltra le richieste e le risposte dell'uno all'altro
 - Il client si collega al proxy, invece che al server, e gli invia delle richieste
 - Il proxy, a sua volta, si collega al server e inoltra la richiesta del client, riceve la risposta e la inoltra al client

19

Perché usare un proxy

- Privacy
 - Se usiamo un proxy quando ci colleghiamo a un sito, il sito vedrà la connessione provenire dal proxy e non da noi
- Aggirare i blocchi di reti private
 - Se non posso parlare direttamente con Tizio, allora lascerò un messaggio a Caio da girare a Tizio
- Aggirare la censura governativa
 - Basta collegarsi tramite un proxy che si trova in un altro Paese

20

Tipi di proxy

- Siti internet gratuiti come
 - vhide.info
 - proxify.com
 - hidemyass.com
 - anonym.to
- Software da installare sul PC
 - TOR (permette di cambiare identità virtuale ogni volta che lo si desidera)
 - Not My Ip (permette di cambiare l'indirizzo IP)

21

Proxy e ricerca personalizzata

- La navigazione anonima permette di sfuggire alla personalizzazione
 - Non essendo riconoscibili nemmeno dal motore di ricerca
 - Facendo finta di essere in un luogo/Paese diverso (quello dove si trova il proxy)
- Possiamo usare i proxy per mettere a confronto ricerca personalizzata e non
 - Eseguiamo la stessa ricerca con il nostro account e tramite proxy
 - Il confronto dei risultati ci fa capire dove e come agisce la personalizzazione

22

Ricerca personalizzata e ricerca tramite proxy: esempi

23

Es. 1: query “Ristorante” - w/ proxy

Search results for "ristorante" (Web, Immagini, Maps, Shopping, Notizie, Altro, Strumenti di ricerca).

Circa 28.900.000 risultati (0,26 secondi)

Ristoranti: Verona: Leggi recensioni sui ristoranti di ... - TripAdv...
www.tripadvisor.it/Restaurants-g187871-Verona_Province_of_Verona_...
Ristoranti a Verona, Provincia di Verona: su TripAdvisor trovi 32.043 recensioni di 793 ristoranti a Verona, raggruppati per tipo di cucina, prezzo, località e altro.
La Cantina del 15 - Antica Torretta - Trattoria dal Taio - La Griglia

Ristoranti Michelin Verona : la Guida Michelin
www.viamichelin.it/web/Ristoranti/Ristoranti-Verona_-Verona-Italia
Trova velocemente un ristorante a Verona grazie a ViaMichelin. I migliori ristoranti della selezione Michelin a Verona organizzati in base al prezzo e alla cucina.

Guida Ristoranti Verona
www.verona.com/it/ristoranti-verona/
Guida aggiornata e dettagliata dei ristoranti a Verona. Trova il tuo ristorante a Verona o consulta il nostro elenco ristoranti Verona.

IL DESCO Ristorante
www.ildesco.com
4 recensioni Google

Trattoria Pane e Vino
www.trattoriapanevino.it
3,7 ★★★★★ 9 recensioni Google

Ristorante 12 Apostoli
www.12apostoli.com
4,6 ★★★★★ 20 recensioni Google

Trattoria Al Pompiere Srl
alpompiere.tv
4,4 ★★★★★ 22 recensioni Google

Al Bersagliere
www.trattoriaalbersagliere.it
4,2 ★★★★★ 12 recensioni Google

Antica Bottega del Vino S. R. L.
www.bottegapvini.it
4,4 ★★★★★ 21 recensioni Google

Mappe per ristorante

Personalizzazione: ristoranti in Verona

- A Via Dietro San Sebastiano, 7 Verona 045 595358
- B Via Giuseppe Garibaldi, 16 Verona 045 800 8261
- C Vicolo Corticella San Marco, 3 Verona 045 596999
- D Vicolo Regina D'Ungheria, 5 Verona 045 803 0537
- E Via Dietro Pallone, 1 Verona 045 800 4824
- F Via Scudo di Francia, 3 Verona 045 800 4824

24

Es. 1: query “Ristorante” - w/o proxy

The screenshot shows a Google search interface with the query 'ristorante' in the search bar. The URL bar displays 'https://www.google.it/search?&output=search&client=...'. Below the search bar, there are tabs for 'Web', 'Immagini', 'Maps', 'Shopping', 'Notizie', 'Altro', and 'Strumenti di ricerca'. The search results indicate 'Circa 27.200.000 risultati (0,58 secondi)'. The first result is 'Ristorante - Wikipedia' with the URL 'it.wikipedia.org/wiki/Ristorante'. The snippet describes a ristorante as a structure where food and drinks are served. Other results include TripAdvisor links for restaurants in Torino, Rimini, Firenze, and Milano, each with a brief description and a link to read reviews.

Google non riconosce
più dove sono:
manca la geolocalizzazione

25

Es. 2: query “linux” - w/ account

The screenshot shows a Google search interface with the query 'linux' in the search bar. The URL bar displays 'https://www.google.it/search?&output=search&client=...'. Below the search bar, there are tabs for 'Web', 'Immagini', 'Maps', 'Shopping', 'Notizie', 'Altro', and 'Strumenti di ricerca'. The search results indicate 'Circa 214.000.000 risultati (0,22 secondi)'. The first result is 'Kubuntu | Friendly Computing' with the URL 'www.kubuntu.org/'. The snippet describes KDE Software and Linux. Other results include 'Linux.it | Linux: Cos'è, Chi è, Perché', 'Linux - Wikipedia', 'Notizie relative a linux', 'Homepage | Ubuntu Italia', and 'Ubuntu: The world's most popular free OS'. Each result includes a brief description and a link to read more.

Personalizzazione:
uso di pagine già visitate

26

Es. 2: query “linux” - w/o account

Web Immagini Maps Shopping Notizie Altro ▾ Strumenti di ricerca

Circa 211.000.000 risultati (0,33 secondi)

[Linux.it | Linux: Cos'è, Chi è, Perché](#)

[www.linux.it/](#) ▾

Che cos'è **Linux**, cos'è il software libero, e informazioni utili sulla community italiana.

[Cos'è Linux](#) - [La Partecipazione](#) - [La Community](#) - [Contatto](#)

[Linux - Wikipedia](#)

[it.wikipedia.org/wiki/Linux](#) ▾

Linux (/ˈlɪnuks/, pronuncia inglese [ˈlɪnuks]) è una famiglia di sistemi operativi di tipo Unix-like, rilasciati sotto varie possibili distribuzioni, aventi la ...

[Ubuntu](#) - [Fedora](#) - [Unix-like](#) - [Linus Torvalds](#)

[Notizie relative a linux](#)

[La backdoor Linux si nasconde](#)

Punto Informatico - 4 giorni fa

Symantec individua un agente patogeno per **Linux** particolarmente astuto nel camuffare il traffico malevolo, una minaccia progettata per colpire ...

[Homepage | Ubuntu Italia](#)

[www.ubuntu-it.org/](#) ▾

Comunità italiana di supporto alla distribuzione di questo sistema basato su GNU/Linux. Forum e informazioni per l'utilizzo.

[Download](#) - [Tour online di Ubuntu](#) - [Derivate](#) - [Scopri Ubuntu](#)

[Linux.org](#)

[www.linux.org/](#) ▾ [Traduci questa pagina](#)

Linux Forum. ... The **Linux** Kernel: Configuring the Kernel Part 18. By DevynCJohnson. Sunday. 888 Views · [slide.jpg](#) · [Partition Tables](#). By DevynCJohnson.

[Perché passare a Linux - Majorana](#)

[www.istitutomajorana.it/passare-linux/](#) ▾

Ci sono almeno 27 validi motivi per innamorarsi di **Linux**. Vediamoli insieme. [Indice generale](#).

[Linux | HTML.it](#)

[www.html.it/linux/](#) ▾

Risorse per **Linux**, notizie dal web, articoli e guide, download della settimana.

[Immagini relative a linux](#) - Segnala immagini non appropriate

27

Es. 3: query “informatica” - w/ account

Web Immagini Maps Shopping Notizie Altro ▾ Strumenti di ricerca

Circa 43.600.000 risultati (0,25 secondi)

[Informatica - Wikipedia](#)

[it.wikipedia.org/wiki/Informatica](#) ▾

L'**informatica** (termine che deriva dalla lingua francese informatique, contrazione di informazione automatica) è la scienza che ha per oggetto lo studio dei ...

[Portale:Informatica](#) - [Storia dell'informatica](#) - [Informatica \(università\)](#) - [Programma](#)

[Pagina iniziale-Dip. Informatica-Università degli Studi di Verona](#)

[www.di.univr.it/?lang=it](#) ▾

A2 E B1 - A.A. 2013/2014 - Alessandra Maria Bossi - Bando per gli insegnamenti ancora scoperti presso il Dipartimento di **Informatica** per l'A.A. 2013-2014 ...

[Notizie relative a informatica](#)

[Difesa: Ue, presentare strategia sicurezza informatica nel 2014](#)

Agenzia di Stampa Asca - 2 ore fa

Difesa: Ue, presentare strategia sicurezza **informatica** nel 2014.

[Websense e le minacce informatiche](#)

La Stampa - 4 giorni fa

[ABS Computers srl](#)

[www.abscomputers.it](#)

4,7 ★★★★★ 5 recensioni Google

[Format2](#)

[www.format2.it](#)

4,9 ★★★★★ 10 recensioni Google

[Furlanis Elettronica S.a.s.](#)

[www.furel.it](#)

[Pagina Google+](#)

[Visualizza i risultati per informatica su una mappa »](#)

[Informatica Verona | PagineGialle.it](#)

[verona.paginegialle.it/veneto/verona/Informatica.html](#) ▾

Informatica Verona | Con PagineGialle trovi aziende, negozi e professionisti nella tua città. Cerca **Informatica** Verona.

[Lavoro informatica, sistemista, programmatore Verona su Bakeca](#)

[verona.bakeca.it/informatica-telecomunicazioni-0](#) ▾

25+ elementi - Bakeca: Sito di annunci gratuiti che ti aiuta a trovare lavoro.

Lavoro per casa azienda operante nel settore Informatico ricerca informatico ...

Philmark Informatica ricerca per immediata assunzione un Programmatore Java



28

Es. 3: query “informatica” - w/o account

Web Immagini Maps Shopping Notizie Altro ▾ Strumenti di ricerca

Circa 48.600.000 risultati (0,35 secondi)

Informatica - Wikipedia
it.wikipedia.org/wiki/Informatica ▾
L'**informatica** (termine che deriva dalla lingua francese informatique, contrazione di informazione automatica) è la scienza che ha per oggetto lo studio dei ...

Portale: Informatica
Il Portale Informatica è un punto di raccolta per le voci riguardanti l' ...

Categoria: Informatica
Voci riguardanti l'informatica. Molte voci afferenti a questa categoria ...

[Altri risultati in wikipedia.org »](#)

Notizie relative a informatica
Difesa: Ue, presentare strategia sicurezza informatica nel 2014
Agenzia di Stampa Asca - 17 ore fa
Difesa: Ue, presentare strategia sicurezza **informatica** nel 2014.

Informatica - The Data Integration Company
www.informatica.com/it/ ▾
Informatica ha concluso il pilota in modo brillante, senza richiedere supporto, nei tempi indicati e superando i requisiti che avevamo posto come base per la ...

[Immagini relative a informatica](#) - Segnala immagini non appropriate

Informatica.it
www.informatica.it/ ▾
Informatica.it è un portale verticale sull'Information Technology che mette a disposizione una serie di servizi come dizionari inglesi, dizionari italiani, glossari, ...

dib - Dipartimento di Informatica - Bari
www.di.uniba.it/ ▾
Università degli Studi di Bari A. Moro, Dipartimento di **Informatica** Via E. Orabona, 4 - 70125 Bari tel. +39-080-5441111 - fax +39-080-5442031 C.F.: ...

Dipartimento Informatica - Sapienza - Università di Roma
www.di.uniroma1.it/ ▾
Sito ufficiale del Dipartimento. Informazioni sul Dipartimento: il suo organico, la

Informatica
L'informatica è la scienza che ha per oggetto lo studio dei fondamenti teorici dell'informazione, della sua computazione a livello logico e delle tecniche pratiche per la sua implementazione e applicazione ... Wikipedia

[Segnala errori](#) [Ulteriori informazioni](#)

Visualizza risultati relativi a
Informatica Corporation
Informatica Corporation è il principale fornitore indipendente di software per l'integrazione dei dati. ...

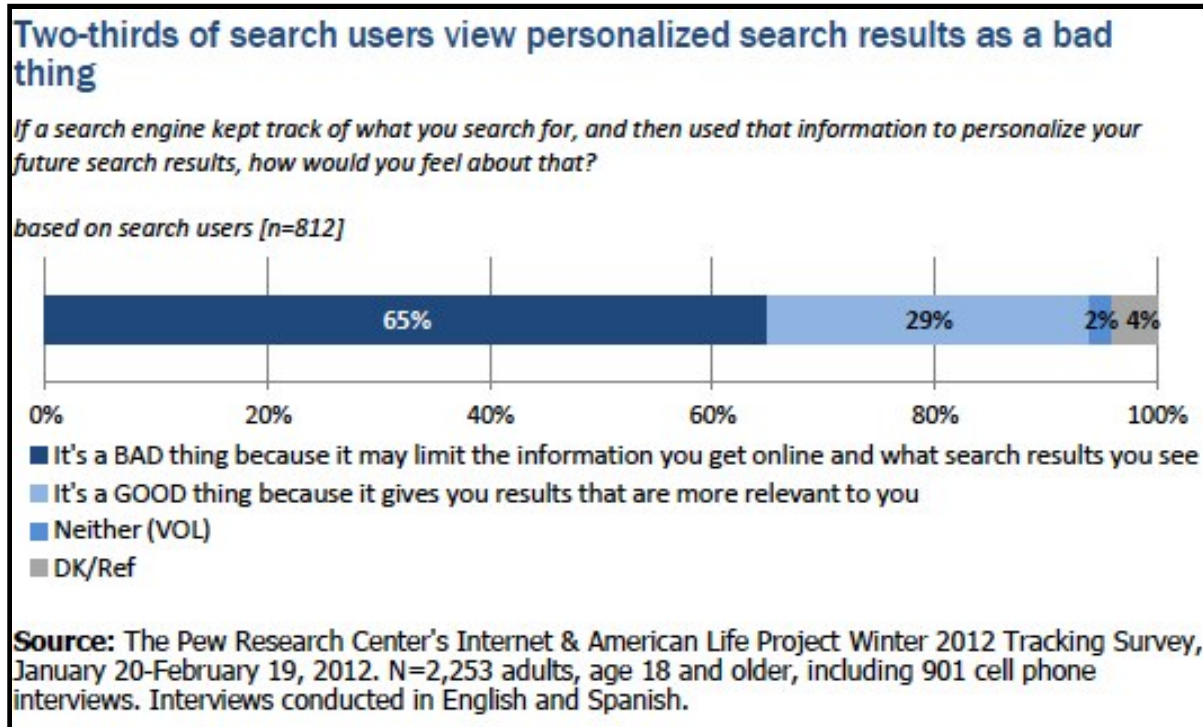
29

Considerazioni

- Ad entrare sicuramente ed evidentemente in gioco sono quindi:
 - Geolocalizzazione
 - Cronologia web
- Più in generale alcuni risultati posso essere ricondotti ai propri interessi

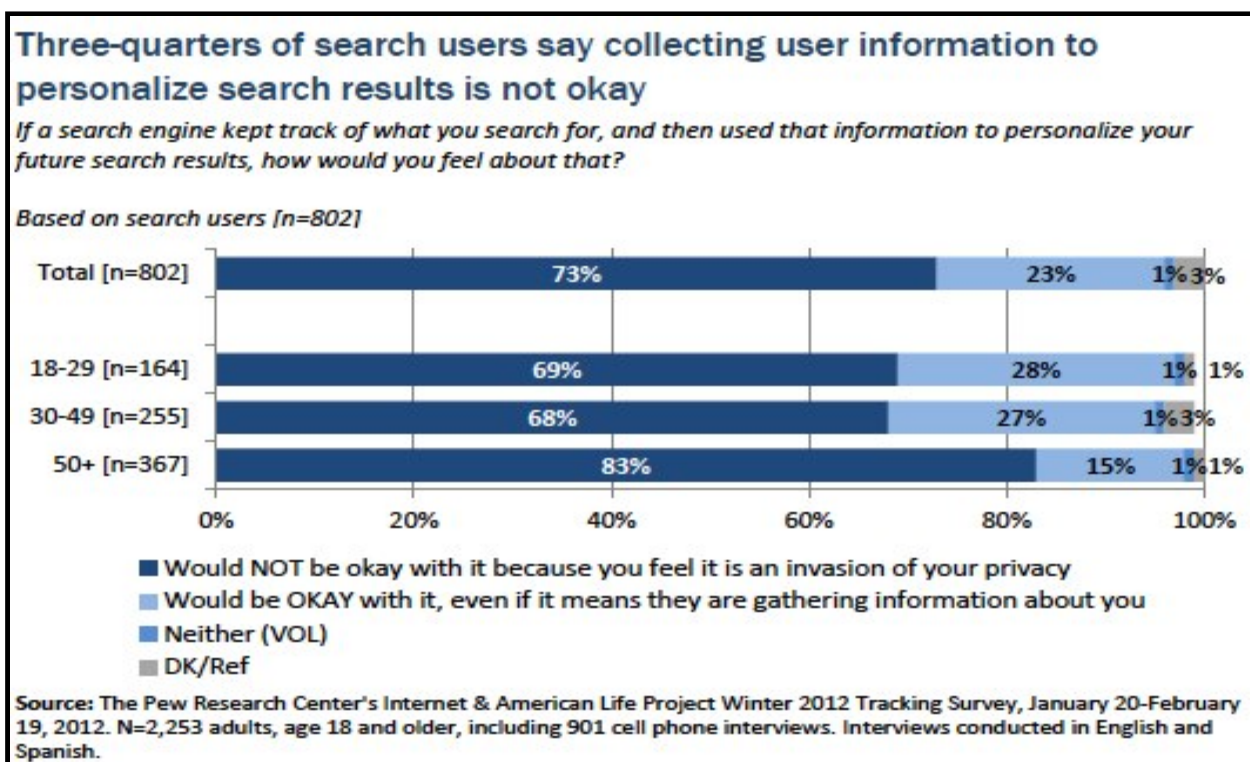
30

Ricerca personalizzata: piace o no? (I)



31

Ricerca personalizzata: piace o no? (II)



Al 73% degli utenti NON PIACE!!

32

Perché?

- Limita le informazioni a cui avere accesso e risultati di ricerca oggettivi
- È percepita come un'invasione di privacy



PARADOSSO:

La ricerca personalizzata contribuisce a migliorare le performance dei motori di ricerca MA è percepita negativamente

33

Come risolvere questo paradosso?

- Gli utenti CONSAPEVOLI dei meccanismi di ricerca personalizzata si proclamano “contrari” ad essa MA prediligono gli effetti positivi di rilevanza e di qualità agli effetti negativi di invasione di privacy e di non-neutralità dei risultati
- Gli utenti NON CONSAPEVOLI ignorano i meccanismi di personalizzazione, per cui non possono che valutare positivamente la ricerca personalizzata

34