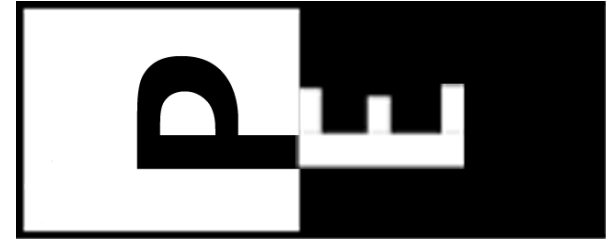


E-privacy 2006

Firenze, 20 maggio



Privacy e motori di ricerca

La minaccia fantasma

Claudio Agosti - vecna@sikurezza.org - <http://www.s0ftpj.org>

Copyright 2006, Claudio Agosti - s0ftpj/sikurezza.org

È garantito il permesso di copiare,
distribuire e/o modificare questo documento
seguendo i termini della GNU General Public
License, Versione 2 od ogni versione successiva
pubblicata dalla Free Software Foundation.

Una copia della licenza è acclusa come nota a
questa slide, ed è anche reperibile all'URL

<http://fly.cnuce.cnr.it/gnu/doc.it/gpl.it.html>

Perché “preoccuparsi” dei motori di ricerca ?

- Da loro parte la maggioranza delle navigazioni
- Sono emersi i più efficienti, eliminando gli altri
- Diamo loro informazioni sensibilissime, più utilizzabili di quanto crediamo
- Centralizzano le informazioni

I dubbi (*finora*) noti

- Profilazione

- Dare le proprie ricerche, come molti altri, per molti anni, significa fornire di log di informazioni sensibilissime

- Informazione guidata

- Quanto conta veramente il ranking ?

- Inclusione indiscriminata nell'indice

- E se non volessi essere indicizzato ?

- La parzialità della rete indicizzata

- Un motore non è l'indice di internet, quali pro e quali contro ?
- Le possibilità di data mining sull'indice, in cosa consistono ?

Precisamente di chi parliamo ?

- Indipendentemente da chi siano i principali motori di ricerca, dallo stato in cui sono, dai servizi che offrono e dalla loro storia, il problema è strutturale
- Google, Yahoo, MSN/Passport, oggi, sono semplici “riferimenti”: non capi d’accusa, non gli unici potenziali problemi

La profilazione

- La profilazione degli utenti è la descrizione digitale in relazione ad un comportamento reale:
 - Es: Il numero di SMS e di e-mail giornaliere
 - Es: I tratti autostradali che effettuiamo e con quale cadenza
 - I nostri interessi, espressi nelle nostre ricerche
- La profilazione delle nostre ricerche descrive, quasi, ogni nostro interesse

La profilazione, 2

- Per mantenere uno storico è necessario potersi ricondurre all'utente
 - Sistemi di tracciamento adeguati e permessi dal protocollo
 - L'utente si fa riconoscere tramite login/password per accedere ad accattivanti servizi
- E' più efficiente correlando comportamenti differenti
 - Nonostante le licenze dicano tutto e nulla

Profilazione - La realtà

- Il cookie eterno
 - Il cookie di google scade il 17/01/2038
- L'interazione tra i servizi
 - Gmail, ads, gruppi, search history
- L'offerta di servizi che motivano la conservazione e l'analisi dei dati
 - Licenze che mirano a dire effettivamente cosa verrà fatto, senza che gli utenti abbiano conoscenza reale del significato

Profilazione - La realtà, 2

- La storia delle nostre ricerche e la teoria dell'”impronta”

mencalgooglebeckysharp@yahoo.co.id | [Home page di Google](#) | [Account personale](#) | [Esci](#)



Cronologia ricerche (versione beta) per mencalgooglebeckysharp@yahoo.co.id

Cronologia ricerche

- [Web](#)
- [Immagini](#)
- [News](#)
- [Froogle](#)

[Seleziona tutto](#)

[Sospendi](#)

[Rimuovi elementi](#)

[Tendenze](#)

[Segnalibri](#) ★

[Aggiungi segnalibro](#)

2 Mag 2006

"ldap explorer" "active directory"

☆ [Ldap explorer software by Softerra LLC and others](#) - 16:04
www.freedownloadscenter.com/Best/ldap-explorer.html

☆ [LDAP Explorer](#) - 16:04
www.asp-dev.com/main.asp?page=88

ldap explorer

☆ [LDAPExplorerTool : an Open Source LDAP Browser](#) - 2 visite - 16:00
ldaptool.sourceforge.net/

milano interval

☆ [Untitled Document](#) - 15:57
web.tiscali.it/interval/

ldap explorer

☆ [LDAPExplorerTool : an Open Source LDAP Browser](#) - 2 visite - 15:12
ldaptool.sourceforge.net/

Attività di ricerca

« Apr Mag 2006

L	M	M	G	V	S	D
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

Oggi, 6 Mag

1 - 5 6 - 10 11 - 20 21+

Totale ricerche: 100

Profilazione - La realtà, 3

- L'uso esplicito degli ads
 - Il data mining applicato alle nostre preferenze diventa pubblicità guidata
- L'uso implicito degli ads
 - Quando ne carichiamo semplicemente uno, che succede ?
- Una nuova forma di analisi
 - Non più “cosa facciamo”, ma “cosa ci interessa”, “cosa ci interessava” e “cosa sappiamo”
- Sfruttamento dell'interattività

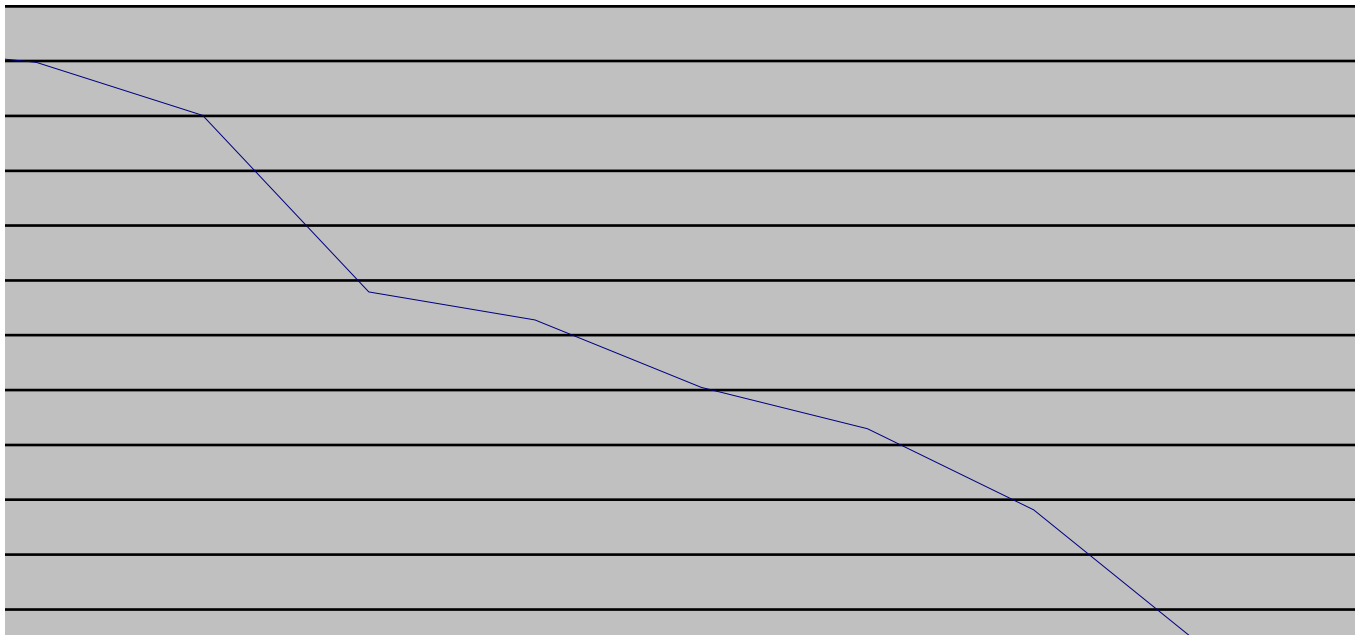
Informazione guidata

- Quando la censura diventa implicita anzichè esplicita, è sempre censura ?
- L'importanza commerciale ?
 - E se l'ordine dell'indice cessa d'essere proporzionale al contenuto, ma è proporzionale all'investimento ?
 - Anche il malware ha come target le prime posizioni

Informazione guidata, 2

- Quanto è importante il ranking ?

Click ricevuti in relazione al ranking



Informazione guidata, 3

- Quando il ranking viene sovvertito ?
 - Attaccando il PageRank, robot e link circolari
 - Pagando, ADS o “collegamenti sponsorizzati”
 - Organizzandosi tramite un SEO
 - Strategia di marketing

- Quando l'informazione perde imparzialità ?
 - Quando il riferimento diventa unico
 - Quando si prevede il funzionamento dei filtri

Indicizzazione di massa

- **Acquisizione implicita di informazioni**
 - Archiviazione virtualmente infinita
- **La prima generazione di blog/liste**
 - Archivi personali online dopo anni possono infastidire i soggetti indicizzati
 - Rottura dei collegamenti di fiducia
- **Non discriminazione dei contenuti**
 - Sollevamento di responsabilità da parte dei SE
 - Prima denuncia su lucro pedopornografico ad un SE

Indicizzazione e servizi limitrofi

- Il lato “privato” del database
 - Le possibilità di correlazione tra servizi, e-mail, dati
 - Analisi automatiche sul contenuto delle e-mail
- Ma è davvero privato ?
 - Le licenze alle quali sottoscriviamo come ci tutelano ?
 - Lo stato in cui il nostro traffico è centralizzato, quale regolamentazione ha a riguardo ?

Indicizzazione, le prime problematiche

- Indicizzare materiale protetto da ©
 - Conservazione di materiale non esplicitamente consentito
 - Fuga di informazioni

- Impossibilità di verifica dei contenuti
 - Attacchi al sistema di ranking
 - Abuso dell'indicizzazione per favorire attività illecite
 - Abusi (Fake SE, acquisizione tramite referrer ?)

La parzialità della rete indicizzata

- Web: rete ad “invarianza di scala”
 - Chi non vorrebbe essere indicizzato può esserlo
 - Chi potrebbe essere indicizzato può non esserlo
 - Velocità di crescita del web e velocità dei crawler
 - Unidirezionalità dei link

Natura del problema

- Problema asimmetrico
 - E' necessario un sistema centralizzato, ma non ci si può affidare solo a quello
 - Prima dei SE l'informazione era di responsabilità dell'informatore
 - La conoscenza dei lettori guida le loro scelte e viceversa
 - Le informazioni che dà la massa sono riconducibili ai singoli
 - Affidarsi a poche entità significa, alla lunga, dare stabilità e potere alle entità in questione

Possibili soluzioni ?

- Qualche idea...
 - Meta SE
 - Ranking pesato in parte dagli utenti (Collaborative Filtering)
 - Proxy di ricerca (<http://www.scroogle.org/>)

Fine

Grazie dell'attenzione