



UNIVERSITÀ
DEGLI STUDI
DI TERAMO



SEO

Search Engine Optimization

-1-

Prof. Romina Eramo

University of Teramo

Department of Communication Sciences

ramo@unite.it

Acknowledgement: Alessandro Celi (University of L'Aquila)

Introduzione al SEO

- **SEO** è l'attività di **ottimizzazione del sito per i motori di ricerca**, ovvero l'applicazione delle tecniche, degli strumenti e delle conoscenze per rendere un sito internet meglio *digeribile* ai motori di ricerca e migliorarne il posizionamento in merito a determinate parole chiave.
- **La SEO non è una scienza esatta**, a causa della variabilità delle tecnologie di ricerca.

Introduzione al SEO

I metodi **SEO** possono essere suddivisi in 3 ambiti principali:

- **SEO sul linguaggio**
- **SEO sull'architettura**
- **SEO sulla reputazione**

E in un sito servono tutti.

SEO sul linguaggio

- **Esprimersi con un linguaggio adeguato al target di utenza a cui ci rivolgiamo**
 - è difficile che un utente possa effettuare una ricerca utilizzando i nostri slogan
 - Bisogna entrare nella testa dei nostri possibili utenti/clienti e scrivere sul sito ciò che loro vogliono trovare.

SEO sull'architettura

Cosa vuol dire?

- **Garantire che le proprie pagine web possano essere indicizzate in maniera semplice**
- **Garantire una adeguata struttura di link per guidare gli spider attraverso i contenuti**

Un crawler (detto anche spider o robot), è un software che analizza i contenuti di una rete (o di un database) in un modo metodico e automatizzato, in genere per conto di un motore di ricerca. Un crawler è un tipo di bot (programma o script che automatizza delle operazioni).

SEO sulla reputazione

- I motori di ricerca pongono grande attenzione alla reputazione dei siti.
- I siti considerati autorevoli avranno un buon posizionamento nei risultati dei motori di ricerca.
- **Per conquistare una buona reputazione bisogna essere linkati da altri siti.**
- **Non tutti i link sono uguali:** quelli che arrivano da siti con una buona reputazione valgono di più.

SEO e Google

- Anche Google ha dedicato spazio alla definizione di SEO

<http://www.google.com/support/webmasters/bin/answer.py?hl=it&answer=35291>

Ha pubblicato anche una guida alla tecnica SEO:

<http://www.google.com/webmasters/docs/search-engine-optimization-starter-guide.pdf>

I Motori di Ricerca

- Partiamo con un classico esempio: l'elenco telefonico cartaceo



PIEVE DI SOLIGO 788

DIEFFE - RAPPRESENTANZE PER L'ARREDAMENTO DI DE FAVERI A. & C. s.a.s.	0438 980 787
56 v. Guglielmo Marconi	
DIEME Lucien Jean Philippe	0438 842 835
17/f v. Antonio Meucci	
DIESSE s.r.l. Imbottiti per Ufficio	0438 83 058
103 v. Montello	
103 v. Montello (n. 2 Linee Urbane)	0438 842 314
DI GENNARO Nicola	0438 980 002
17/s v. John F. Kennedy	
DI GIOIA Anna, 79 bg. Stolfi	0438 841 437
DI GIORGIO Elisabetta Maria Grazia	0438 82 730
1 v. Caldella	
DINI Angelo, 79 v. Brandolini - Solighetto	0438 82 517
	0438 841 050

Libro con un elenco di nominativi inseriti in ordine alfabetico e classificato per località.

Cosa facciamo per cercare un numero di nostro interesse?

I Motori di Ricerca

- Altro esempio è il **classico dizionario**.
Ci sono i termini, con le relative definizioni, in ordine alfabetico.

Per cercare un termine dobbiamo sfogliare il libro, seguendo l'ordine giusto delle lettere fino ad arrivare alla 'parola' di nostro interesse.

I Motori di Ricerca

- Quando l'archivio è molto grande, o i dati da cercare sono tanti, la ricerca è indubbiamente lenta e complessa.
- Per risolvere questo problema esistono i **motori di ricerca**.

Cosa sono i Motori di Ricerca?

- Un motore di ricerca è un sistema automatico che immagazzina una grande quantità di dati ed effettua una classificazione di questi dati in base a formule matematiche, rendendo semplice, veloce ed efficiente la ricerca di informazioni all'interno di questo grande archivio.

Perchè esistono i Motori di Ricerca?

- I motori di ricerca esistono **per risolvere problemi di ricerca di informazioni**, soprattutto quando i dati archiviati sono in numero elevato, oppure sono tali per cui la ricerca alfabetica, alfanumerica e sequenziale sarebbe troppo complessa.

Motori di Ricerca per il Web

- Si tratta di **grandi archivi di dati, che contengono informazioni dettagliate su un gran numero di pagine Web e che permettono una rapida ricerca.**
- **Precisazione:** nei motori di ricerca sono memorizzati singoli documenti che compongono un sito Web. E' possibile quindi che un determinato sito Web sia presente numerose volte nell'archivio, in relazione al numero di pagine di cui esso è composto.

In Google: *site:libero.it* -> pagine indicizzate

Motori di Ricerca per il Web

- **L'inserimento delle pagine Web negli archivi dei motori di ricerca, può avvenire in due modi:**
 - **manuale** (attraverso la segnalazione dell'utente)
 - **automatico** (attraverso un particolare software che riesce a visitare milioni di siti Web al giorno, inserendo nuove pagine ed aggiornando quelle già presenti in archivio).

Il lavoro impossibile dei motori di ricerca

- Cercare di censire tutte le pagine web esistenti in rete è un'impresa impossibile!
- Perché è impossibile?

Il lavoro impossibile dei motori di ricerca

- **Perché il web è frammentato:** possiamo immaginarlo come una foresta frastagliata composta da agglomerati di alberi contigui, ma anche da piante isolate sul cucuzzolo di una montagna o su una impervia vallata, sconosciute ed irraggiungibili, a meno di una segnalazione specifica della loro posizione esatta
- **Perché è in continua evoluzione e crescita:** ogni giorno vengono creati e messi online milioni di documenti nuovi, oppure vengono aggiornati quelli già presenti in rete

La Scienza delle Reti

- Un po' di Storia.

Tratto dal Libro:

Link, La scienza delle reti - Albert-Làszlò Barnabàsi

Einaudi Editore

Il WWW (World Wide Web) è la più grande rete mai costruita dall'uomo:

un insieme vastissimo di documenti, contenuti multimediali e di servizi che possono essere resi disponibili dagli stessi utenti Internet.

II WWW

- **Il Web è una rete virtuale dove i nodi sono le pagine web che contengono qualunque tipo di informazione: documenti, immagini, video, ...**
- **Il potere del web risiede tutto nei Link o URL (Uniform Resource Locators), grazie ai quali è possibile spostarsi tra risorse web collegate.**

Quanto è grande il Web?

- Quanti documenti contiene? Quanti link?
Fino a qualche anno fa nessuno lo sapeva, nemmeno approssimativamente.
- A raccogliere questa sfida furono 2 ricercatori del NEC.
Secondo la loro rilevazione, già nel 1999 il web conteneva 1 miliardo di documenti.



3,333,179,849

Internet Users in the world



1,005,264,807

Total number of Websites



72,297,038,289

Emails sent **today**

Quanto è grande il Web?

- **Il fattore che conta non è la grandezza del Web, ma la distanza da una risorsa ad un'altra.**
Quanti click sono necessari per spostarsi da un documento ad un altro, anche se si trova dall'altra parte dell'Oceano?
- Se le pagine web si trovano a migliaia di click di distanza l'una dall'altra, senza un motore di ricerca non potremmo mai rintracciare un documento.

I connettori

- Esiste quindi una manciata di persone che hanno la capacità di stringere un numero eccezionale di amicizie chiamati **‘Connettori’**.
- I connettori hanno una grandissima importanza nella nostra società. Sono coloro che reggono i fili di interconnessione delle relazioni sociali.
- Si tratta di **nodi con un numero incredibilmente alto di link.**

La loro scoperta ha rivoluzionato tutte le conoscenze già acquisite sulle reti.

Il web è realmente democratico?

- Se così fosse tutto ciò che è pubblicato nel web avrebbe la stessa opportunità di essere letto da milioni di utenti.
- Se effettivamente la legge che regola questa rete fosse casuale probabilmente sarebbe vero. Ma la realtà è ben diversa.

Il web è realmente democratico?

- Una volta pubblicate le nostre idee nel web, tutti hanno la possibilità di leggerle
- Diventano accessibili a chiunque, in qualunque parte del mondo, con una semplice connessione ad internet.
- **La domanda giusta da porsi è un'altra.**

Qualcuno mi noterà nel web?

- **Se pubblico qualcosa in rete qualcuno la noterà?**
- **Per essere letti bisogna essere visibili:** una banale verità che vale tanto per gli scrittori quanto per gli scienziati. **Sul web la misura della visibilità sono i link: il numero di link e il peso di ogni link.**

Qualcuno mi noterà nel web?

- Ogni pagina web ha in media 5-6 link che puntano a una delle miliardi di pagine web esistenti. **La probabilità che il creatore di un documento, inserisca un link alla mia pagina web è prossimo allo zero.**
- Se per esempio la homepage del mio sito web è linkata da 50 pagine web presenti in rete, ed il web è composto da 10 miliardi di pagine web, la probabilità che la mia pagina venga visitata è di $50/10.000.000.000 \sim 0,0000005\%$.

Qualcuno mi noterà nel web?

- Dalla mappatura del web portata avanti dai ricercatori del Notre Dame University, su un campione di 200 milioni di pagine web, era uscita fuori una caratteristica molto interessante: **il 90% delle pagine web non riceveva più di 3 link dall'esterno.**
C'erano invece 2 o 3 pagine web che ne ricevevano addirittura milioni.
- Parliamo dei Portali e dei Motori di Ricerca: Yahoo, Altavista, Amazon.

Gli Hub

- Così come avviene nella società umana, pochi individui conoscono un numero insolitamente alto di persone (che svolgono il ruolo di *connettore*), **l'architettura del WWW è dominata da pochissimi nodi altamente connessi denominati Hub. E' chiaro quindi che tutti i nodi poco conosciuti, scarsamente visibili e dotati di un esiguo numero di link sono tenuti insieme da questi rari siti altamente connessi.**

Gli Hub

- **Collettivamente creiamo in qualche modo degli hub.**
Si tratta dei siti a cui tutti si collegano, facilissimi da trovare, rintracciabili da qualsiasi punto, tipicamente i motori di ricerca.
- Rimangono però dei quesiti:
Come si formano gli hub?
Quanti ne può accogliere una data rete?

Legge di Potenza

- In rete esistono molti nodi con pochissimi link e pochissimi nodi con moltissimi link (hub).
- Il risultato: la distribuzione dei link su varie pagine web seguiva una ben precisa espressione matematica, detta legge di scala o legge di potenza.

Il fattore crescita

- **Le reti nel mondo reale tendono ad avere una caratteristica comune: la crescita.**
Tutte le reti che conosciamo sono nate da una manciata di nodi per poi estendersi sempre più: il Web è un esempio lampante.

I ricchi diventano sempre più ricchi

- La crescita, da sola, non riesce ancora a spiegare la presenza di hub o connettori e non possiamo ancora spiegare quindi le leggi di potenza.

Collegamento Preferenziale

- **I nodi più connessi vincono!**
- Chiunque, quasi inconsciamente, tende a connettersi con nodi che già conosce, quasi sempre con quelli più connessi del web.

Si tende a preferire gli Hub. Dovendo scegliere tra due pagine tendiamo a visitare quella che riceve il doppio dei link, insomma quella più connessa.

Collegamento Preferenziale

- L'evoluzione della rete è governata dalla legge del collegamento preferenziale: tendiamo inconsciamente ad aggiungere link ai nodi già super collegati.

Perché nascono gli Hub?

- Il ruolo più importante è svolto dalla crescita: i primi nodi hanno più tempo per acquisire link, rispetto ai nodi più giovani. **La crescita offre un chiaro vantaggio ai nodi più vecchi che diventeranno i più ricchi di link.**
- **Con il collegamento preferenziale inoltre i nuovi nodi preferiscono connettersi con i nodi più ricchi di link.**

Perché nascono gli Hub?

- Con l'arrivo di nuovi nodi che continuano a scegliere i nodi più connessi, i nodi più antichi accumuleranno un numero altissimo di link, distanziandosi dal gruppo, fino a formare degli Hub.
- Possiamo dire che la velocità con cui ogni nuovo nodo attrae nuovi link è proporzionale al numero di link che già possiede.

Perché nascono gli Hub?

- Ci sono comunque degli aspetti delle reti nel mondo reale che non sono spiegabili tramite le leggi di potenza ed il modello ad invarianza di scala:
 - i link interni tra vecchi nodi
 - la sostituzione di un link con un altro
 - la rimozione improvvisa di link
 - l'invecchiamento di un nodo con conseguente perdita di interesse
 - etc...

Come fanno i nuovi a sopravvivere?

- **Come fanno i nuovi arrivati a sopravvivere in questo spietato mondo?**
- In quasi tutti i sistemi complessi, ogni nodo ha caratteristiche speciali che vanno al di là del suo grado di connessione. Le pagine web, le aziende e gli attori hanno proprietà intrinseche che influenzano la velocità con cui, in un ambiente competitivo, riesco ad attrarre nuovi link.

Come fanno i nuovi a sopravvivere?

- Alcuni nodi, pur comparando molto tardi, si aggiudicano in breve tempo la maggior parte dei link della rete.
Altri pur arrivando molto presto non se ne aggiudicano nessuno.
- Caso InkTomi, partner di Yahoo, sostituito dal nuovo arrivato Google...
- Caso De Havilland annientato dopo appena un anno dall'inizio dei voli dalla Boing. Boing stessa che molti anni dopo vedette perdere un'importante quota di mercato a vantaggio della Airbus.
- Possiamo affermare quindi che i nodi non sono tutti uguali.

La competizione nei sistemi complessi: modello a fitness

- Benchè sia impossibile trovare la chiave universale del successo, possiamo studiare il processo che separa i vinti dai vincitori: la competizione nei sistemi complessi.
- Introduciamo quindi **la fitness: la capacità competitiva di ogni nodo.**

La capacità di una pagina Web di farci tornare quotidianamente sul suo contenuto, anziché su quello di altri miliardi di pagine che si contendono la nostra attenzione.

La competizione nei sistemi complessi: modello a fitness

- **Tra due nodi che hanno la medesima fitness, quello più vecchio rimane comunque il favorito nella gara di link.**

La competizione nei sistemi complessi: modello a fitness

- Come ha fatto Google a diventare un Hub da un giorno ad un altro?
- Ultimo arrivato, ma con una grande tecnologia alle spalle come motore di ricerca, conquistò nuovi link molto più rapidamente dei suoi concorrenti, fino a surclassarli.
La “bellezza” vince sull’età.

Chi vince prende tutto

- **E' possibile che in alcune reti 'chi vince prende tutto'.**

Il comportamento di una rete non dipende dalla natura dei suoi nodi e dei suoi link, ma da come è distribuita la sua fitness.

Un esempio? Microsoft Windows che ha conquistato praticamente l'intero mercato, pur non essendo il primo sistema operativo inventato.

Chi vince prende tutto

- **I nodi sono in competizione continua perché i link rappresentano la principale fonte di sopravvivenza in un mondo interconnesso.**

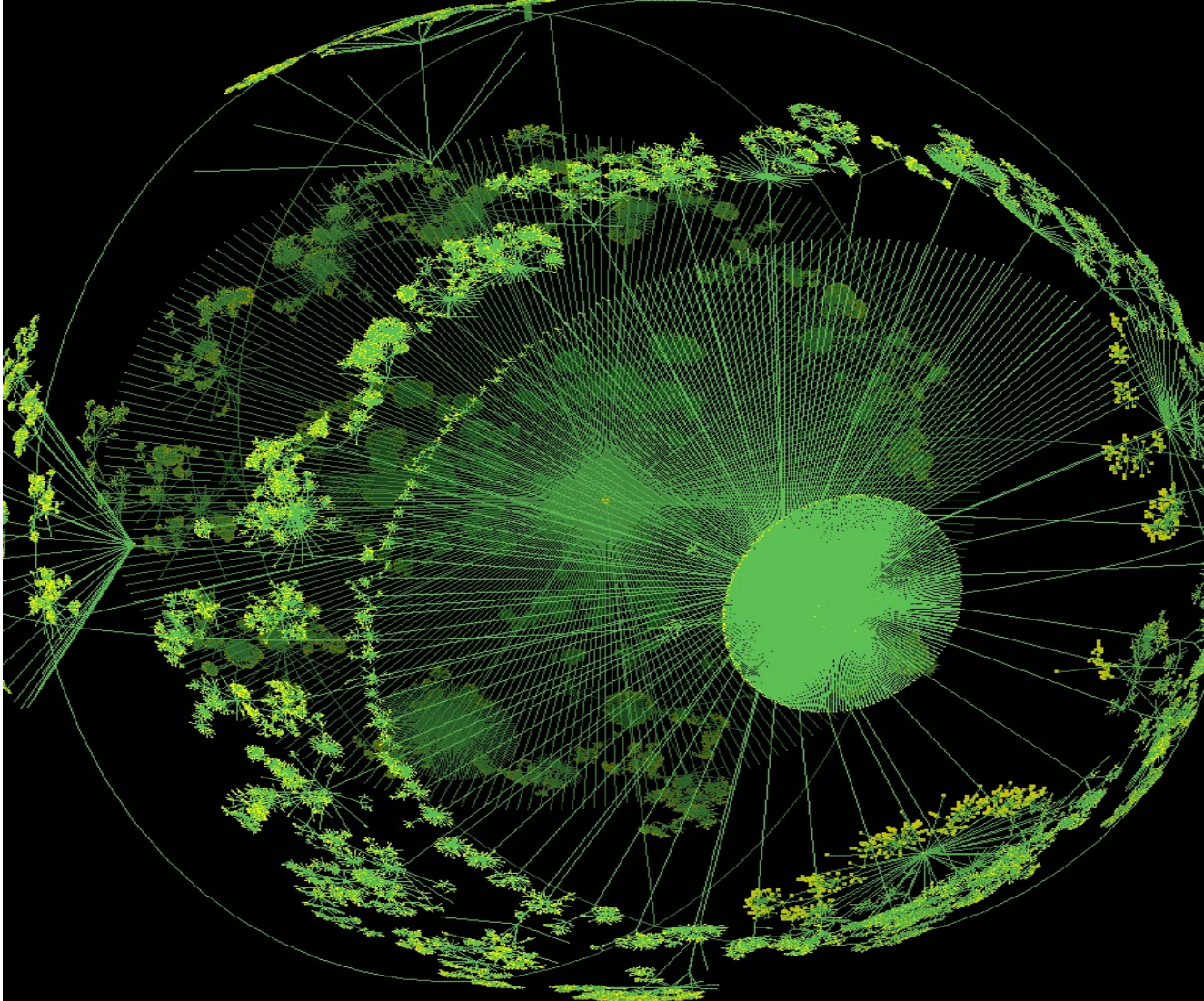
Finché continueremo a scegliere dei nodi e a scartarne altri ci saranno sempre dei Vincitori e dei Vinti.

Si tratta di sistemi competitivi dove i nodi lottano con forza per ottenere nuovi link.

Internet

- Capire la topologia di Internet è il requisito fondamentale per progettare strumenti e servizi che offrano un'infrastruttura comunicativa veloce e affidabile.
- In internet esiste una stretta relazione tra la densità dei nodi e la densità di popolazione. Nuovi nodi nascono dove c'è più richiesta, ossia dove c'è un certo numero di persone interessate.

Internet



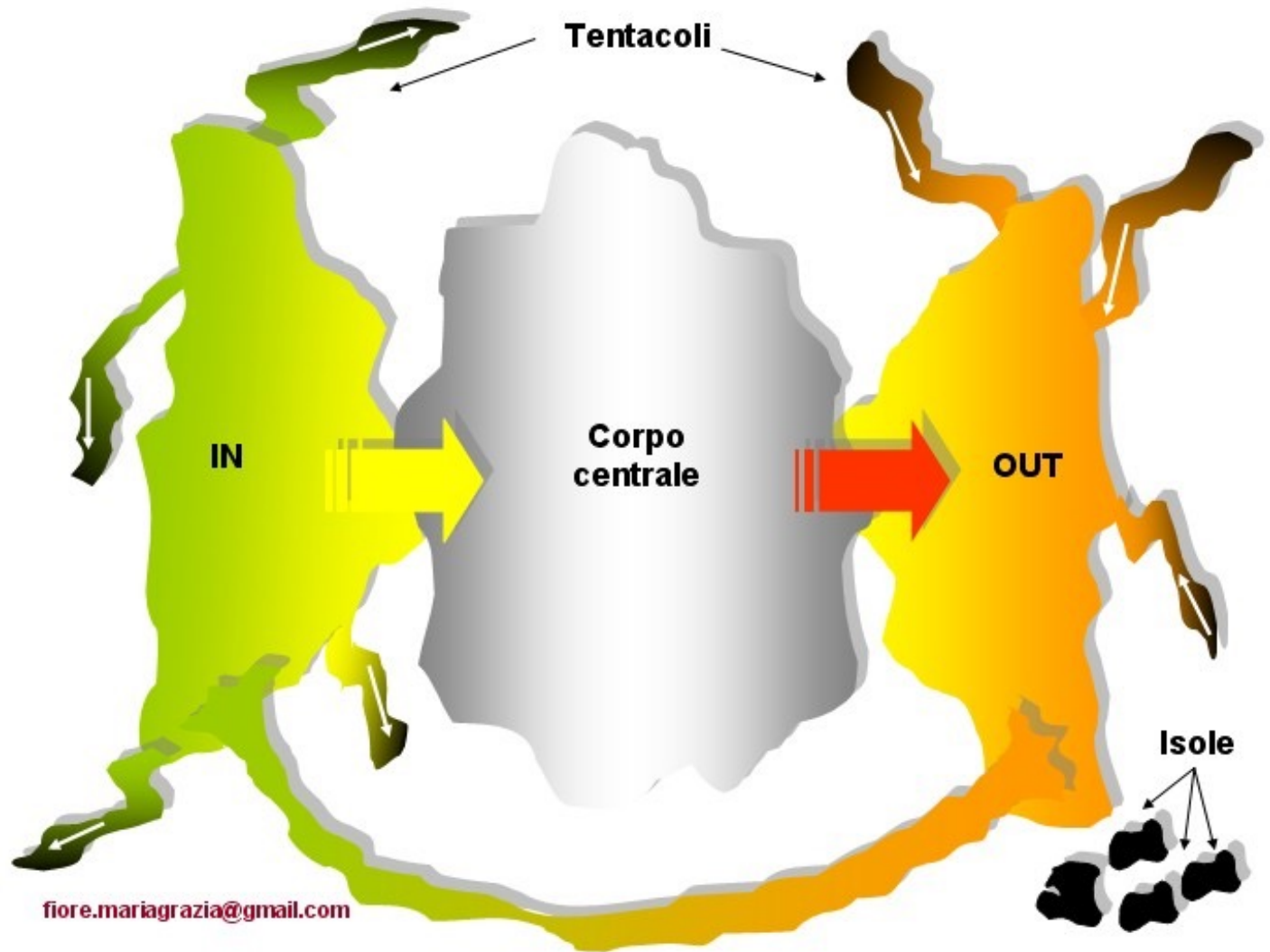
Il Web Frammentato

- Cosa implica il web frammentato e perché è così?
- - **Difficoltà per gli spider di indicizzare tutto il web.**
- - **Difficoltà nel restituire i risultati più pertinenti.**
Sarebbero più utili 20 risultati eccellenti che 100.000 di dubbia utilità.
- - **I link del web sono <<orientati>>:** lungo un dato url si può viaggiare in un'unica direzione.

Il Web Frammentato

- Il web avendo link diretti non forma una singola rete omogenea.
- E' diviso in **4 grandi continenti**, ciascuno con le sue regole di navigazione (scoperta di Andrei Broder di Altavista con altri collaboratori di IBM e Compaq, 1999).

I Continenti



I Continenti: Corpo Centrale

- Il primo di questi continenti comprende **un quarto di tutte le pagine web**. Chiamato *corpo centrale* ospita **tutti i più grandi siti web**. **Altamente connesso, di facile navigazione**. Questo non vuol dire che tutti i documenti sono connessi a tutti gli altri, ma semplicemente che è **molto semplice in quest'area trovare un percorso per raggiungere due nodi a caso**.

I Continenti: IN e OUT

- Il secondo e terzo continente, chiamati IN e OUT, hanno le stesse dimensioni del corpo centrale ma sono più difficili da navigare. **Dalle pagine del *continente IN* è semplice raggiungere il corpo centrale, ma una volta lì non ci sono percorsi per tornare indietro.**

I Continenti: IN e OUT

- Viceversa, i nodi appartenenti al *continente OUT* sono facilmente raggiungibili dal corpo centrale, ma non ci sono link per tornare indietro. Possiamo vedere per esempio il continente OUT come una serie di siti aziendali che possono essere facilmente raggiunti, ma una volta entrati non si può tornare indietro.

I Continenti: Tentacoli e Isole

- Il quarto continente è formato da *tentacoli* ed *isole*: **gruppi di pagine connesse tra loro ma non raggiungibili dal corpo centrale.** Gruppi di pagine connesse tra loro che possono portare dal continente IN al continente OUT, e viceversa, oppure rimanere completamente inaccessibili a chi non ne conosca l'indirizzo preciso.
- Circa $\frac{1}{4}$ delle pagine web si trovano in questo continente.

I Continenti

- **E' importante capire che la posizione di una pagina nel web dipende, più che dal suo contenuto, dalle sue relazioni con gli altri documenti, tramite i suoi link dall'esterno e quelli all'interno.**

Se la vostra pagina si trova in un isola, i motori di ricerca non la scopriranno mai, a meno che non segnalate l'URL manualmente.

I Continenti

- La possibilità quindi di tracciare tutto il web per un motore di ricerca non è solo una questione di algoritmi, tecnologie e risorse economiche. **I motori di ricerca riescono a mappare solo la parte centrale ed il continente OUT, quindi il 50%. Il resto rimane nascosto.**

Per quanti sforzi i robot o spider possano fare non riusciranno mai a trovare i documenti lì situati.

I Continenti

- **Con l'aumentare della grandezza del Web in un futuro si riuscirà a formare un unico continente che assorba tutti e 4?**

La risposta è semplice: **NO!**

- **Finché i link saranno diretti questo non si verificherà mai.** E i link nel web non diventeranno mai bilaterali



I Continenti

- I continenti non sono un'esclusiva del World Wide Web.
- Anche la rete degli articoli scientifici con le loro citazioni hanno una direzionalità e quindi hanno anch'essi i 4 continenti.
- Anche la rete alimentare ha una direzione: il leone mangia l'antilope, ma l'antilope non mangia il leone.
- **Tutte le reti dirette sono divise negli stessi 4 continenti.**

Le Comunità o Social Network

- I 4 continenti comunque non sono le uniche aree definibili nel web.
- Su scala più piccola, al loro interno, proliferano piccole cittadine e metropoli. Sono i siti che, accomunati da un'idea, un hobby o un habitat, creano delle comunità di interessi condivisi: gli appassionati di musica rock o di carpfishing o di opera lirica. Possiamo chiamarle **comunità o social network**.

Le Comunità o Social Network

- **Ogni volta che creiamo un link verso un'altra pagina Web, ne facciamo risaltare l'importanza rispetto alla nostra area di interesse.** Quindi presumibilmente i link di un appassionato di calcio a 5 possono condurci ad altri dello stesso tipo, permettendoci di ricostruire la comunità di appassionati di calcio a 5.

L'individuazione di queste comunità virtuali ha un enorme potenziale applicativo: sociale, economico, organizzativo.

Le Comunità o Social Network

- **Sorge un problema: tra miliardi di pagine web, è possibile identificare queste comunità?**
- **I ricercatori NEC avevano formulato la loro ipotesi: appartengono ad una stessa comunità i documenti che hanno più link fra loro che link rivolti all'esterno della comunità.**

Le Comunità o Social Network

- La deduzione dei ricercatori NEC era un ottimo punto di partenza per realizzare algoritmi che identificassero diversi raggruppamenti all'interno della topologia del World Wide Web.
- Anni fa questa cosa sarebbe stata impossibile.
- Google invece ha dimostrato che ciò è assolutamente possibile, costoso algoritmicamente, ma possibile.

Le Comunità o Social Network

- Il **web** è quindi **suddiviso** in vari **continenti**, ognuno dei quali accoglie vari villaggi e territori che hanno la forma di **comunità sovrapposte**. **Perché sovrapposte?**

Una comunità che parla di letteratura inglese potrebbe essere identificata in più raggruppamenti o sottoinsiemi:

siti in inglese, siti letterari, siti dedicati a Shakespeare, etc...

Democrazia in Rete

- Abbiamo visto che nella topologia ad invarianza di scala i documenti hanno una visibilità mediamente scarsa, perché una minoranza di siti altamente popolari detiene quasi tutti i link.
- **Nel web c'è la libertà di parola, ma le nostre voci possono essere troppo deboli per farsi sentire.**

Democrazia in Rete

- **Le pagine con pochi link dall'esterno non verranno mai individuate da una ricerca casuale, perché gli hub tendono a catalizzare tutta l'attenzione su di loro.**
- Sembra impossibile per i robot sfuggire a questa trappola della popolarità.
- Le pagine che hanno un solo link dall'esterno hanno il 10% di possibilità di essere indicizzate. Quelle che ricevono tra 21 e 100 link dall'esterno hanno invece il 90% di possibilità di venire indicizzate.

Democrazia in Rete

- **L'evenienza che qualcuno trovi la mia pagina Web dipende solo da un unico fattore: la mia posizione nella rete.**
- Se tanti visitatori si connettono alla mia pagina web, la trovano interessante, e creano un link dalle loro pagine, allora diventerà pian piano un hub, che i motori di ricerca dovranno notare per forza.
- Se la mia pagina web viene ignorata, finirà nell'oblio delle pagine dimenticate e isolate del web.