

Comportamento degli utenti

• Legge di Fitts

La legge di Fitts viene usata per calcolare il tempo impiegato per puntare un obiettivo (che può essere inteso come un'area o un bottone), partendo da una certa posizione e terminando sopra di esso, in base alla distanza dell'oggetto e alla dimensione di questo oggetto. In altre parole, più grande e vicino è l'oggetto e più veloce sarà lo spostamento sopra di questo. La formula è così espressa: $T = a + b \cdot \log_2 \left(1 + \frac{D}{W}\right)$ dove

- T: è il tempo medio impiegato per concludere il movimento;
- a: è il tempo costante impiegato dall'essere umano per cominciare a concludere l'azione;
- b: è la costante che dipende dagli strumenti utilizzati e dall'utente;
- D: distanza dal punto iniziale alla zona obiettivo;
- W: ampiezza della zona obiettivo.

L'applicazione più efficiente di questa legge si verifica quando con il cursore attiviamo un menù contestuale: il punto più rapido in cui si trova è quello in cui c'è già il puntatore. Altre aree "magiche" sono i bordi e gli angoli. Per ulteriori approfondimenti, leggere qui: <http://www.usabile.it/272004.htm>

Visibilità

• Cos'è lo spam index?

Lo spamdex è l'insieme di attività svolte con lo scopo di far sì che un sito sia indicizzato dal motore di ricerca e occupi le posizioni più alte della SERP (Search Engine Results Page). Lo spam index è conosciuto anche come SEO (Search Engine Optimization) o SEP (Search Engine Positioning). Quando si effettua una ricerca, ad ogni pagina viene assegnato un punteggio, punteggio calcolato in base a due componenti: quella testuale (data dal TFIDF) e da quella ipertestuale. Il TFIDF (Term Frequency Inverse Document frequency) è il risultato del prodotto del TF per il IDF. Per calcolare il TF, si calcola la percentuale di occorrenza di una parola in una pagina, per esempio se ci troviamo davanti a una pagina web formata da 100 parole e ricerchiamo la parola "pippo" che ricorre 5 volte, avremo un TF pari al 5%. Risulta ovvio che articoli e preposizioni avranno un TF maggiore rispetto ad altre parole poichè sono di collegamento e non di contenuto. L'IDF, invece, calcola l'inverso della frequenza della parola nell'insieme dei nostri documenti in scala logaritmica. Quindi, per esempio, su 1000 pagine l'articolo "il" appare in 980 pagine e risulta un TF=98% mentre il IDF= $\log(1/0,98\%)=0,008$. Invece, se su 1000 parole la parola "pippo" ha il 10% di frequenza il suo IDF sarà uguale a 1. Risulta chiaro come più rara è la parola e più è alto il bonus. Dando una definizione semplificata: un motore di ricerca prende tutte le pagine contenenti la parola cercata e calcola di ognuna il TFIDF, se si ricercano più parole, si sommano i TFIDF di tutte le parole ricercate.

- **Tecniche per aumentare il posizionamento**

Per aumentare il posizionamento, si deve cercare di aumentare il TFIDF delle parole chiave del sito. Per fare ciò, esistono varie tecniche sul dove posizionare queste keywords:

- **Body spam:** si inseriscono le parole nel body della pagina HTML, è efficiente ma svantaggia la lettura per l'utente;
- **Title spam:** si inseriscono le parole chiave nei titoli della pagina (tag title), il contenuto viene modificato molto meno;
- **Meta tag spam:** le keywords vengono inserite in appositi tag html (i meta, appunto) che descrivono a keyword la pagina. Il vantaggio principale è che il contenuto non viene modificato, ma i motori di ricerca odierni li ignorano per l'intenso sfruttamento.
- **Anchor text spam:** le parole chiavi vengono inserite nel testo dei link e il punteggio assegnato rispetto alle comuni parole è maggiore, anche la pagina di destinazione riceve dei benefici in termini di punteggio poichè viene aumentato.
- **URL spam:** le parole chiavi si trovano nell'indirizzo della pagina stessa e anche essi ricevono dei punti bonus rispetto a comuni parole nel testo (come per l'anchor text spam).

- **Quali tecniche vengono utilizzate per inserire le keywords**

Le parole chiavi posso venire inserite in diversi modi, per esempio:

- **Repetition:** la keyword viene ripetuta stando attenti al suo TFIDF. Questa tecnica è facile da individuare dai motori di ricerca che possono penalizzare il punteggio della pagina.
- **Dumping:** vengono inseriti moltissimi termini poco usati anche se questi non sono inerenti al testo, risulta efficace contro quelle queries che includono termini relativamente rari e poco chiari: per queste queries è probabile che solo un paio di pagine siano rilevanti, così anche una pagina spam con basso livello di rilevanza/importanza potrà apparire tra i risultati migliori.
- **Weaving:** inserisco nel mio sito pezzi di altri siti web in cui sono inserite le nostre keywords in modo random, creando un contenuto più interessante e valutato positivamente dai motori di ricerca che assegneranno al sito un punteggio più alto in questo modo.
- **Stitching:** viene fatto il past© da siti differenti e si assembla il tutto ottenendo un contenuto rilevante. Più argomenti diversi danno inoltre un bonus globale e i SE fanno fatica a capire se il contenuto è copiato.
- **Broadening:** inseriscono non solo le keywords, ma anche i sinonimi, le parole simili e frasi correlate. Non solo per coprire le richieste degli utenti, ma anche per sfruttare la tecnica di somiglia dei motori di ricerca così da ottenere bonus aggiuntivi.

- **Spiegare il problema e le soluzioni al term spamming**

Usare lo spamdex per l'utente non è molto gradevole e crea disapprovazione. Per ovviare a questo problema, vengono utilizzate varie tecniche:

- **Content hiding:** il contenuto potenziante è nascosto all'utente e solo i motori di ricerca lo vedono.
- **Redirection:** detta anche tecnica 302, consiste nel potenziare una pagina che sarà costituita con soli fini di posizionamento che indirizza alla vera pagina. Così gli utenti vedranno la vera pagina studiata per loro mentre i SE si fermeranno a leggere la prima. Questa tecnica funziona solo con javascript poichè i SE riconoscono il meta che reindirizza mentre utilizzando js il meta verrà aggiunto dinamicamente e i SE non si accorgeranno di questo in quanto non eseguono il codice. C'è da dire che i SE odierni riconoscono questa tecnica, quindi per ingannarli è sufficiente offuscare un po' il codice.
- **Cloaking:** è una specie di redirection che controlla se la pagina è richiesta da un utente o da un motore di ricerca e a seconda della risposta ricevuta viene presentata una pagina apposita per il bot e una per l'utente. È una tecnica difficile da scoprire e per questo se il motore di ricerca scopre che si sta utilizzando questa tecnica, si viene bannati per un periodo.

- **Cos'è il pagerank**

Per calcolare il punteggio di una pagina, viene valutato il punteggio relativo al pagerank: ovvero l'algoritmo di pagerank assegna un peso numerico ad ogni elemento di un collegamento ipertestuale presente nel sito. Per calcolare questo valore si utilizza la formula: $\pi_v = \sum_{(w,v) \in E} \left(\frac{\pi_w}{d_w} \right)$ dove

- w, v è un link da w a v ;
- d_w è il numero di link uscenti;

Quindi, all'aumentare del numero dei link complessivi dei siti che puntano alla nostra pagina, il pagerank aumenta e lo si può vedere come il numero delle probabilità di entrare nel nostro sito. C'è da tenere in considerazione anche il fattore "random surfer": si fanno navigare questi random surfer, chiamati anche "spider", che raccolgono dati su cui calcolare il pagerank. Esistono però dei problemi:

- **Spider taps:** nel caso in cui uno spider rimanere intrappolato dentro un sito in cui le pagine continuano a indirizzare ad altre pagine in un ciclo infinito (esempio: calendario)
- **Island:** ci sono casi in cui alcuni pezzi di rete rimangono isolati dagli altri e gli spider non hanno alcun modo per passare, quindi si generano infiniti valori diversi di pagerank oppure nessun valore di pagerank.

Per risolvere a questi problemi, la formula è stata modificata aggiungendo il "teletrasporto": $\pi_v = (1 - \epsilon) \cdot \left(\sum_{(w,v) \in E} \left(\frac{\pi_w}{d_w} \right) \right) + \left(\frac{\epsilon}{N} \right)$ In pratica si è aumentata la casualità e ora il random surfer ha anche un dado e con una certa probabilità effettua il teletrasporto in un'altra pagina. Nonostante ciò, questa formula non risolve il problema delle dead ends.

- **Come migliorare il pagerank?** Per migliorare il pagerank con l'uso del teletrasporto, ci sono due fattori importanti: inlink e outlink. Infatti, per

aumentare il pagerank di una pagina dobbiamo fare in modo di aumentare i link entranti, ovvero avere un elevato numero di siti che puntano a quella pagina. Per fare ciò, esistono diverse tecniche:

- Infiltration: infiltrarsi in altri siti e mettere link alla pagina da potenziare;
- Honey pot: creare contenuto appetibile e utile che riceve link (modo giusto di aumentare il pagerank);
- Link exchange: allearsi con altri siti per scambiarsi link;
- Resurrection: acquistare domini defunti con pagerank molto alto, infatti il pagerank non muore mai e rimane collegato al dominio, oltre al fatto che l'età del dominio conta come bonus

Per quanto riguarda gli outlink, quando viene posizionato un link uscente dalla pagina un po' di flusso di pagerank esce dalla pagina, ma ciò non significa che non possa aumentare il pagerank della mia pagina. Infatti posso utilizzare molte tecniche per aumentare il pagerank, sia utilizzando gli inlink che gli outlink.

- **Spam farm**

La spam farm è una tecnica che viene utilizzata per aumentare il pagerank di un sito. Con spam farm si indicano le strutture che raccolgono link a pagine web spam permettendo così l'aumento di pagerank di altri siti che risulteranno avere un elevato numero di link entranti. Una spam farm ottimale consiste in una pagina o in più pagine, pagine target, puntate da link bidirezionali puntate da altre pagine, pagine potenzianti. Il modo migliore per sfruttare le spam farm è farlo tramite alleanze con altri siti. Posso avere:

- Alleanze profonde, in cui le pagine target distribuiscono il flusso sulle pagine potenzianti dell'altro sito (tutto punta a tutto, link bidirezionali);
- Alleanza superficiale, le pagine target dei due siti sono connesse con link;
- Alleanze ring, alleanza che comprende più di due siti, le pagine target indicizzano in un percorso circolare la pagina target successiva creando una sorta di anello;
- Alleanza complete core, il potenziamento è simile a quello dell'alleanza ring, ma possiede una struttura bidirezionale che la rende più vantaggiosa. Le pagine target hanno link entrate e uscenti verso le due pagine target più vicine.

Web semantico

- **Scrivere cos'è, pro e contro dell'RDF**

Il modello RDF (Resource Description Framework) è un framework comune che descrive metadati, relazioni e concetti interoperabili, ovvero permette lo scambio di informazioni senza necessità di traduzioni in linguaggi. Si basa su una grammatica di base semplice: soggetto-predicato-complemento

oggetto. Questi tre elementi possono contenere due diversi tipi di dati: URI oppure stringhe letterali, che compongono la struttura linguistica e logica più semplice. L'RDF può essere rappresentato come un grafo e può essere rappresentato in due differenti modi: attraverso XML utilizzando la struttura per alzare la complessità e arricchire il significato oppure attraverso le N-TRIPLE utilizzando le triplette "soggetto-predicato-comp. oggetto". Quest'ultimo è molto più semplice dato che è stato appositamente fatto per RDF. Per integrare RDF con il web, si è utilizzato XHTML2 con cui viene introdotto RDFa che aggiunge gli attributi about(il soggetto) e property(il verbo) mentre il contenuto è il complemento oggetto. Usando RDF, dove quindi le informazioni vengono rappresentate utilizzando i grafi, funziona molto bene l'aggregazione: unendo due grafi di conoscenza si genera un grafo creato tramite collegamenti automatici dei nomi del web(URI). Grazie a questa proprietà, RDF risulta migliore rispetto ai dialetti XML. Funzionano bene anche la reificazione, ovvero si riduce ad un oggetto una asserzione, ottenendo una diversificazione di livelli e garantendo la monotonicità, cioè se ci si fida dell'insieme allora sono giusti anche i singoli pezzi che lo contengono; i contenitori, concetti logici di AND e OR; le variabili, oggetti logici non specificati; e la] monotonicità, come detto precedentemente, preso un grafo, se l'informazione espresso in esso è vera allora tutti i sotto-grafi sono veri. Inoltre, più struttura permette una maggiore integrità e una maggiore deduzione: se ho una classe che contiene un oggetto specifico, si può dedurre che questo oggetto specifico è di tale classe anche se non è esplicitamente scritto.

- **RDF-schema**

L'RDF-schema è lo standard che permette il supporto di ontologie, per dare informazioni più semplici vado quindi a fare una classificazione diminuendo così la complessità. La struttura informativa degli oggetti è data da: class, subclassOf e individual(concetto di individuo), mentre per i verbi abbiamo a disposizione property(la relazione, il verbo), subPropertyOf, Domain(tipo di relazione che delinea il verbo), range(dove la proprietà è applicata). Grazie a questi strumenti possiamo dare informazione e categorizzarla

- **FOAF: descrizione, ambiti d'uso, pregi e difetti**

Il progetto FOAF (Friend Of A Friend) è un'ontologia utilizzata per definire un vocabolario RDF al fine di descrivere i metadati relativi alle persone e sui loro interessi, relazioni e attività. Questa iniziativa è nata con l'obiettivo di creare una rete di dati processabili. Come i documenti HTML, anche i documenti FOAF possono essere collegati tra loro per formare una rete di dati con una semantica ben definita. Risulta quindi chiaro che la sua caratteristica principale è quindi quella di definire le relazioni tra le persone, grazie alla relazione foaf:knows, per esempio, posso esprimere il concetto di amicizia e collegare più persone. Difetti: attualmente si sta lavorando per risolvere i problemi di affidabilità delle informazioni, garantire e avere la possibilità di correggere eventuali errori nel caso in cui alcuni dati siano sbagliati e infine distinguere i dati pubblici da quelli privati.