

Accesso

Necessita' di indicizzazione

"...le carte coprivano gia' la meta' del pavimento. -Si e' fatto molto lavoro - disse il sindaco crollando il capo - e quella non e' che una piccola frazione. Il piu' l'ho riposto nel granaio, e veramente la massima parte e' andata perduta. Come si fa a conservare tutto? Ma nel granaio c'e' ancora molta roba. Potrai trovare il decreto? - disse rivolgendosi alla moglie - Devi cercare un documento in cui la parola agrimensore e' sottolineata con la matita azzurra - ..."

Cosi' Kafka nel suo "Il Castello" (come altri grandi letterati di un passato ancora del tutto analogico⁷¹) anticipava non solo le odierne paure di non riuscire ad accumulare e catalogare "tutta" l'informazione, ma addirittura alludeva con l'accenno alla "sottolineatura azzurra" alla necessita' di indicizzarla.

L'indicizzazione e' infatti una necessita' che occorre quando terminata la

fase di recupero dell'informazione, bisogna catalogare e classificare i dati, pronti per essere successivamente interrogati.

Selezionata la porzione di Rete da recuperare (e quindi, conseguentemente esclusa tutta la Rete che non si è voluto o potuto scandagliare), la prima scelta⁷² importante che il motore di ricerca effettua avviene sulla porzione di documento da indicizzare: Altavista ed Hotbot per esempio indicizzano solo i primi 2-300 caratteri del documento, mentre Arianna indicizza il testo intero.

Non ci stancheremo mai di far notare come in maniera diversa ma articolata, e secondo procedure molto rigorose, tutti i motori di ricerca provvedano a ri-produrre - e allo stesso tempo a deformare l'informazione. E se il diritto alla deformazione dell'informazione⁷³, così come alla sua produzione, è un principio stabilito dalle prime comunità hackers nelle quali ci riconosciamo⁷⁴, è altrettanto vero che tale deformazione (o produzione) dovrebbe avvenire secondo procedure rese note all'utenza finale, interagibili con essa e comunque basate su meccanismi che non siano facilmente manipolabili dalle esigenze di mercato o di controllo/censura dell'informazione.

Chiusa la riflessione ideologica, possiamo provare ad addentrarci nei funzionamenti tecnici del processo di indicizzazione di un motore di ricerca⁷⁵.

La fase di indicizzazione consiste nella creazione di indici cosiddetti invertiti, allo scopo di facilitare l'interrogazione. Ma chi indicizza, utilizza le stesse parole di chi effettua l'interrogazione di ricerca sul motore ? E' un problema dai molteplici risvolti dal momento che l'interfaccia utente che svolge la query di ricerca va ad interrogare un database a sua volta indicizzato secondo regole precise che potrebbero anche aver escluso, o comunque non sottolineato, la parola inclusa nel documento, a causa della sua frequenza o collocazione nel documento stesso⁷⁶. Il risultato della ricerca, e soprattutto il criterio di classifica (ranking) dei risultati con cui sarà fornita risposta alla query, dipenderà quindi da queste regole.

La fase di indicizzazione corrisponde ad un processo di analisi dei documenti che andranno a costituire il database. E' in questa fase che le parole portatrici di funzioni sintattiche che non contribuiscono direttamente a specificare un contenuto informativo - le cosiddette stopwords - dovranno essere eliminate per essere distinte dalle parole riconosciute come significative⁷⁷. Dopodiché si giunge alla creazione degli indici invertiti: l'elenco della porzione di testo in cui si trova la parola significativa affiancata da un coefficiente che esprime la sua frequenza nel testo.

Attualmente, la maggior parte dei motori di ricerca utilizza l'algoritmo di Porter insieme ad altri algoritmi che sono in grado di ponderare, oltre alla frequenza, anche variabili qualitative come la posizione delle parole

citare all'interno del testo.

Ricapitolando, il motore di ricerca dell'informazione seleziona a priori una porzione di Rete da cui recuperare l'informazione di ipotetico interesse, dopodiché analizza questa informazione in toto oppure solo in parte (i capoversi iniziali).

L'analisi dell'informazione e la sua catalogazione - l'opera di vera e propria indicizzazione - avviene secondo regole formulate in base ad algoritmi matematici, che cercano di ignorare le parole stimate non significative. L'esclusione dei termini avviene passando attraverso una "griglia linguistica" fondata sull'impiego di dizionari linguistici che ignorano tutte le preposizioni, congiunzioni ecc. - oppure secondo criteri quantitativi, che cioè si riferiscono alla loro frequenza nel testo, troppo alta o troppo bassa. Ciò che rimane delle parole considerate significative viene a questo punto "pesato" a seconda della sua frequenza e collocazione nel testo analizzato, ed è in questa fase che localizzazione e reiterazione dei termini giocano un ruolo essenziale.

Facciamo due esempi pratici. Una parola posta all'interno dei tag h1 78 è sicuramente interpretata come molto importante e dunque spicca agli "occhi" digitali dello spider. Analogamente si verifica per una parola che appare raramente in una pagina, e molto di frequente in un'altra facente parte della stessa documentazione: lo spider infatti suppone di trovarsi di fronte, rispettivamente, a un indice e a un approfondimento dell'indice

stesso e quindi considera il termine di particolare interesse.

Tutto ciò si traduce nel fatto che il motore, una volta formulata un'interrogazione, produrrà un elenco di risposte stilando una classifica.

Una lista che cercherà di soddisfare le esigenze dell'utente che effettua la query presentando i documenti contenenti la parola chiave richiesta in un ordine numerato secondo una graduatoria di rilevanza presunta. In altre parole, la presenza della keyword in posizioni topiche di particolare risalto, non solo nei meta-tags ma anche nel testo stesso, e riportata nella frequenza opportuna, pregiudica il ranking.

Quanto osservato mostra la necessità di produrre un'informazione che sia accessibile al motore di ricerca. Un "tipo di informazione" che possa essere facilmente recuperato e correttamente indicizzato, poiché la visibilità dell'informazione agli spider coincide, giocoforza, con l'accessibilità ai motori di ricerca⁷⁹.

5.2

Forme accessibili di informazione

L'informazione deve essere in primo luogo reperibile, quindi interpretabile.

Una considerazione quasi lapalissiana che tuttavia non viene presa nelle dovute considerazioni dalla maggior parte dei produttori e distributori dell'informazione in Rete.

La preoccupazione che invece accomuna tutti gli autori e i web masters che bramano diffondere informazione in Rete, e' quella di confezionare nella maniera graficamente piu' attraente possibile un oggetto da collocare sull'ambita ragnatela ipertestuale, senza perdere troppo tempo in inutili riflessioni generali.

La smania di poter apparire sulla vetrina piu' luccicante e fantasmagorica che l'uomo abbia mai potuto immaginare fa spesso dimenticare i principi fondamentali per cui l'informazione, per essere effettivamente tale, debba essere reperita, e quindi "letta" e interpretata. In primis, dai motori di ricerca!

Una pagina web per avere qualche chance di visibilita' da parte del motore ha principalmente due possibilita': essere molto popolare oppure farsi notare.

La popolarita' di un documento sotto il profilo informativo si esprime nel fatto di essere linkato da molte altre pagine, alcune delle quali sono gia' inserite nell'elenco delle pagine da visitare. I motori infatti, visitando il web in maniera ipertestuale - ovvero di link in link - sicuramente prima

o poi capiteranno in questa pagina fino ad allora sconosciuta, ma che gode di molti links su pagine già note al motore. La popolarità - intesa nel gergo dei motori di ricerca come quantità di links di cui gode una pagina sul web - ha una forte influenza non solo nel ranking di una pagina ma rappresenta anche una maniera molto efficace per farsi notare dal motore: una sorta di selezione naturale per cui solo i più linkati si fanno notare e prevalgono sugli altri arrivando primi nelle classifiche dei motori.

L'altra modalità classica per farsi notare dai motori di ricerca è quella di segnalare loro la propria esistenza attraverso i comandi add url / add site. Basta perlopiù compilare un semplicissimo form in rete in cui si specifica l'url della pagina ed e-mail di riferimento, e si ottiene l'iscrizione di una pagina all'elenco delle pagine da visitare da parte dello spider⁸⁰. Il sito potrà essere di conseguenza visitato e sottoposto al processo di indicizzazione solo se sarà effettivamente reperibile online.

Anche questa, che appare come una considerazione di sconcertante banalità, nasconde invece delle problematiche tecniche spesso disattese. Un sito per essere reperito facilmente dal motore di ricerca deve risiedere su una macchina in Rete ben configurata e in buona salute in termini di funzionamento e banda. L'indirizzo non deve riscontrare difficoltà in termini di indirizzamento⁸¹; né deve riscontrare problemi di reperimento sul sito che lo linka. Infatti se il sito è linkato troppo in profondità nella macchina⁸², il motore che visita la macchina in Rete deve compiere un

numero troppo elevato di salti di link in link per arrivare al sito partendo dall'homepage della macchina visitata; e, qualora si riscontrino problemi di connettività, questo può ostacolare il reperimento del sito stesso. La macchina in Rete che ospita il sito deve garantire una presenza online costante (cosa mai scontata in un sistema complesso come Internet) e con buone possibilità di banda di connessione (cosa ancor più difficile da riscontrare) in modo tale che il motore, che periodicamente compie la visita, non incontri intoppi di comunicazione, finendo per effettuare un lavoro largamente incompleto di reperimento dei siti previsti. In relazione ai problemi di connessione, si suggerisce di non produrre pagine eccessivamente pesanti in termini di byte: un'informazione asciutta e leggera ha maggiori possibilità di farsi irretire nelle trame dello spider.

Un altro ostacolo sulla via del reperimento è rappresentato dal caso molto comune di cambiare indirizzo alle pagine, senza lasciare traccia del nuovo url nel vecchio. Quante volte accade che gli utenti del vecchio indirizzo - compresi i motori di ricerca - incontrano difficoltà a rintracciare la nuova versione della pagina? Ecco un'altra indicazione che pare ovvia ma viene sistematicamente disattesa dalla maggior parte dei web masters, siti istituzionali compresi!

Ammesso, con buona dose di ottimismo, che la pagina in Rete sia facilmente reperibile dai motori di ricerca, si pone ora il problema di renderla "digeribile" dal motore stesso il cui metabolismo interpretativo segue

regole ed algoritmi differenti dai browsers di navigazione.

Non e' infatti scontato che le pagine web possano essere interpretate correttamente dallo spider del motore di ricerca. Per tentare di comprendere a fondo il problema, immaginiamo che il motore di ricerca legga le pagine di un sito come un browser di vecchia data, una specie di lynx⁸⁴ appositamente configurato per visionare in maniera veloce un'enorme quantita' di materiale.

Allora, tanto per cominciare, molti motori di ricerca non leggono tutte le pagine di un sito, ma prendono in esame solo la parte iniziale ("in alto") delle pagine che visitano. Questo vuol dire che non solo viene "tagliato" tutto cio' che si trova al di sotto di questo confine virtuale ("in basso"), ma significa che la discriminante fra "cio' che e' in alto e cio' che e' in basso" per il motore di ricerca puo' essere anche molto diverso da quello che l'utente vede sullo schermo. Cosa e' "in basso - in alto" nel caso di pagine fatte tramite formattazione a tabelle, molto comuni ultimamente? E nelle pagine a frames che spesso e volentieri non vengono interpretate correttamente dal motore stesso? E nelle pagine che contengono javascript in cima, sebbene ai nostri occhi non visibile? Gli occhi del motore di ricerca non sono i nostri⁸⁵ e i parametri - discriminanti ottiche - non si equivalgono quasi mai. Analogamente succede nel caso in cui lo spider del motore si imbatta in un altro tipo di codice o comandi, relativi ad esempio ai sempre piu' imperanti plug-in ed effetti speciali del Web⁸⁶. Inoltre molti motori di ricerca si trovano in difficolta' quando trovano

comandi automatici di reindirizzamento delle pagine come il comune refresh che spesso viene interpretato come un "trucco" per avvantaggiare il ranking di una pagina. Una volta individuato, la pagina che presenta il refresh viene penalizzata dal motore stesso.

Anche senza eccessive complicazioni tecniche e senza mettere peraltro in dubbio l'effettiva utilità di qualche plug-in alla moda, proponiamo l'analisi di una pagina web composta unicamente da un collage di immagini anche suggestive. Appare sì come una pagina dotata di senso sotto il profilo estetico ed artistico rispetto ad un visitatore umano⁸⁷, ma totalmente priva di significato rispetto a un motore che deve prelevarla e indicizzarla correttamente. Ogni informazione produce/deforma senso a seconda dei media con cui si interrelaziona e dei contesti tecnologici con cui si interfaccia.

Non solo il motore di ricerca può andare in certi casi⁸⁸ in crisi in presenza di oggetti difficilmente interpretabili come immagini senza alternativa testuale, plug-in, codice javascript o di altro tipo, frames, comandi di refresh ecc.; ma è anche vero che il motore può reagire alla presenza di alcune pagine e interpretarle come "illecite", penalizzandole alla stregua di un rigorosissimo cybercensore.

Ma che cosa è una pagina illecita per un motore di ricerca?

Sono considerate illecite dai motori di ricerca tutte quelle pagine che

sono state realizzate appositamente non per offrire un contenuto informativo a chi naviga nel cyberspazio, ma per sottolineare particolari contenuti agli spider. E' il caso delle cosiddette doorways: pagine che ripetono ossessivamente e in maniera considerata efficace rispetto a taluni motori di ricerca, le parole-chiave destinate cosi' a salire nel ranking degli stessi motori. Sono pagine molto curiose per la loro pochezza di contenuti e per la ridondanza della scarsa informazione presente e che hanno la proprieta' di reindirizzare velocemente, una volta analizzate, il motore sulla pagina vera e propria⁸⁹.

Questa, insieme ad altre funzioni⁹⁰ che vengono interpretate da alcuni motori come trucchi escogitati per pompare le pagine in cima alle classifiche, diventa allora causa di penalizzazione da parte di molti motori. Pagine cosi' concepite non vengono neanche prese in considerazione, oppure inserite in una "lista nera" che le valuta con un grado di attenzione minima. Il motore - purché dotato di automatismi di analisi ben configurati - riesce quindi in alcuni casi a distinguere e penalizzare un atteggiamento considerato illecito.

Nell'ipotesi di progettare un sito con l'obiettivo di renderlo subito popolare dentro, e quindi anche fuori⁹¹ del cyberspazio, e cioè realizzare un sito accessibile ai motori di ricerca, e' necessario affrontare alcuni passaggi. Ne elenchiamo per chiarezza i piu' significativi:

- 1) Inventare un nome o un acronimo quanto piu' originale⁹²;
-

2) acquistare il dominio relativo al nome scelto⁹³ oppure farsi ospitare da un dominio in cui ci riconosciamo per condivisione di idee o attività professionali;

3) verificare presso chi ospita il dominio e il sito che sussiste un corretto e costante funzionamento della macchina in Rete in particolare per quanto riguarda le possibilità di banda e connettività;

4) realizzare pagine web originali e creative che soprattutto non abbiano caratteristiche ostili ai motori di ricerca⁹⁴;

5) iscrivere il sito manualmente⁹⁵ ai motori di ricerca ritenuti interessanti;

6) rendere nota - evitando di disturbare pesantemente per non rischiare l'effetto opposto - la nascita del sito attraverso annunci e iniziative sia in Rete che attraverso altri mezzi di comunicazione, così da suscitare interesse e forse ottenere anche qualche benefico link;

7) utilizzare efficacemente tutti gli strumenti leciti per una corretta indicizzazione del materiale informativo contenuto nel sito: l'attributo alt di descrizione delle immagini, il title della pagina, i meta tag (keywords e description), il testo compreso fra i tag h1. In altre parole mettere l'accento su cosa si presume appaia allo spider del motore di

ricerca soprattutto in cima alla pagina, controllando che anche nei nomi dei files, delle dir e del dominio appaia la parola-chiave che vogliamo sottolineare. E così via.⁹⁶

Come si è potuto intuire non è affatto scontato realizzare un sito ottimizzando la sua reperibilità e indicizzazione rispetto ai motori di ricerca.

Ma perché auspicare di essere trovati e ben apparire sui motori?

Perché desideriamo diffondere al maggior numero di persone le nostre informazioni. Lo scopo finale è la visibilità. E dunque, come ci siamo preoccupati* fino ad ora di rendere accessibile il nostro ipotetico sito ai motori di ricerca, dobbiamo come logica conseguenza cercare di renderlo accessibile anche alla totalità della potenziale utenza che frequenta il web⁹⁷.

Come rendere accessibile l'informazione su web è una problematica enorme⁹⁸ e finalmente emergente. Il pressing caparbiamente effettuato sull'argomento in questi anni da personaggi storici della telematica come Arcelele, congiunto all'opera di lobbying da parte di categorie interessate al problema come quelle dei disabili⁹⁹, sta producendo i primi effetti.

Mentre va in stampa questo approfondimento probabilmente sarà apparsa in Parlamento la prima proposta di legge in materia allorché iniziative

simili già esistono in altri paesi (Francia in testa), e i primi segnali di un interessamento politico si sono già avvertiti nell'estate 2000 anche in Italia, in concomitanza con i seminari tenuti all'hackmeeting del CSA101 Forte Prenestino, sulla necessità di rendere universalmente accessibili almeno i siti di interesse pubblico. Gli enti pubblici hanno l'obbligo morale e il dovere civile di realizzare siti progettati secondo principi di accessibilità universale, poiché la loro funzione pubblica non può giustificare una rappresentazione informativa penalizzante nei confronti di una parte dell'utenza.

Ma che cosa significa rendere universalmente accessibile un sito? Rispetto ai motori di ricerca vuol dire renderlo correttamente reperibile e interpretabile da parte di uno spider in cerca di informazioni. Ma accessibilità "universale" significa espandere questo concetto verso un ambizioso progetto che comprenda tutti gli utenti possibili - compresi i soggetti che possono avere difficoltà di varia natura nell'utilizzare strumenti hardware o software.

Infatti se la preoccupazione di un web master - anche perché perlopiù sollecitato in questo lato dal committente del lavoro - è in genere quella di produrre un sito che sia il più colorato, multimediale e intrigante possibile secondo i canoni delle estetiche di moda su Internet, è pur vero che la fruizione universale di materiale informativo sul Web non segue esattamente questa logica. La creazione di un sito che offline appare come un CDROM perfettamente funzionante, ricco di immagini - video - suoni -

plug-in flash e chi piu' ne ha piu' ne metta, e' infatti un'operazione legittima che appaga l'occhio e il senso di meraviglia garantendo il consenso del cliente. Tuttavia questa ricerca estetica dell'effimero non e' affatto detto che assicuri la visibilita' sui motori, ne' tantomeno rappresenta la migliore via di veicolare un'informazione in Rete. Innanzitutto e' una pratica che non ottiene risultati ottimali sui motori i quali incontrano una serie di difficolta' tecniche per reperire e indicizzare il sito. Inoltre non rispetta neppure le esigenze della maggior parte dell'utenza che probabilmente non coincide con un navigatore dotato di un'ottima connessione, hardware e software all'ultimo grido, ma sta spendendo tempo e denaro, abbagliato da effetti speciali che gli allungano i tempi di connessione prima di arrivare al nocciolo dell'informazione cercata.

Ci sono persone che hanno hardware e softare vecchio, di tipo molto diverso oppure diversamente configurato da quello auspicato dal webmaster per visualizzare correttamente le pagine. Ci sono ancora persone che hanno il monitor "settato" a 640x480 e non a 600x800. Ci sono persone con scarse risorse macchina e di banda. Ci sono persone che non amano girare con gli active X, cookies e gli applet Java abilitati perche' gelosi della propria privacy e invulnerabilita' digitale. Ci sono persone che non hanno mai scaricato il plug-in flash oppure quello real-audio anche solo perche' non ne conoscono l'esistenza e diffidano di misteriosi messaggi di invito a scaricarli. Ci sono persone che hanno il tempo contato per girare sul web e quindi non hanno voglia di perdersi in onirici labirinti di effetti

speciali. Ma soprattutto esistono persone che possono addirittura essere danneggiate da alcuni movimenti o frequenze di movimento inaspettate degli oggetti ospitati sulla pagina¹⁰², oppure che non riescono a seguirne consciamente il percorso comunicativo¹⁰³. Esistono persone che non vedono e quindi necessitano di hardware e software di tipo particolare in grado di poter tradurre le pagine web -solo se realizzate secondo alcuni principi di base ispirati all'accessibilita' - in testo parlato o Braille. Esistono persone che non sentono e che quindi necessitano di traduzioni testuali dell'audio presente su un sito.

Ora, se consideriamo nella sua globalita' che Internet non siete voi - il vostro cliente e qualche persona con la quale occasionalmente chattate, ma il web e' un potenziale infinito di cui e' impossibile sapere le caratteristiche, e le problematiche psico-fisiologiche, ne' tantomeno le caratteristiche tecniche di postazione telematica utilizzata per navigare, allora forse non sara' difficile concordare che esiste in Rete un universo umano e tecnologico che attualmente sfugge alle logiche di marketing, ma che non puo' e non deve sfuggire alla nostra coscienza e sensibilita'.

La liberta' di una persona e' la liberta' di tutt*.

Questa parola d'ordine e' una delle tante che cercano di descrivere la complessita' di vivere la solidarieta' in senso universale e che spesso viene solo interpretata nei termini restrittivi di atteggiamento caritatevole. Tuttavia oltre a questa accezione dovrebbe implicare un

significato piu' profondo perche' cio' che e' universalmente accessibile, cio' che e' utile per le persone che hanno meno risorse, e' in realta' utile per tutt*!¹⁰⁴

Realizzare una pagina web accessibile¹⁰⁵ e' un vantaggio per tutt*, di utilita' universale soprattutto in determinate situazioni avare di connettivita'. Costruire una pagina che offra in maniera spartana ma scaricabile velocemente un contenuto informativo chiaro e puntuale, rappresenta una risorsa di grande importanza. Se poi tutto cio' si scontra comunque con necessita' aziendali di realizzare pagine molto pesanti ma altrettanto attraenti, si puo' proporre di progettare una doppia versione del sito: accanto alla versione high-tech, una universalmente accessibile.

Diventano quindi termini di una rivendicazione politica le strategie tecniche¹⁰⁶ da adottare per realizzare una pagina ad accessibilita' universale¹⁰⁷.

Accanto ai criteri di accessibilita' nei confronti dei motori di ricerca, indichiamo altri utili suggerimenti per costruire una pagina accessibile.

- evitare di usare codice proprietario¹⁰⁸ non conforme agli standard del W3C¹⁰⁹; evitare di ottimizzare le pagine solo per una data versione di browser o per una risoluzione video maggiore di 640 x 480¹¹⁰

- cercare di suddividere la pagina in sezioni chiaramente riconoscibili in

modo da ottimizzare la ricerca delle informazioni sul sito; cercare di distinguere fra struttura e presentazione¹¹¹; valutare attentamente i tempi di caricamento delle pagine¹¹² considerando che spesso gli utenti che visitano un sito: sono alla ricerca di qualcosa di preciso, non conoscono la struttura del sito, non hanno molto tempo da dedicare alla visione di ogni pagina, non amano i lunghi tempi di attesa nel download

- per i links impiegare un testo che abbia senso anche al di fuori del contesto e in una lettura che avviene in successione¹¹³

- non scrivere documenti aventi linee che superino le 80 battute per riga; utilizzare valori relativi % e non assoluti per le tabelle¹¹⁴ e gli altri oggetti della pagina, utilizzare ascii-standard¹¹⁵

- garantire un buon contrasto tra testo e sfondo ovvero assicurare che testo e grafica siano percepibili e comprensibili anche se visualizzati senza colori e comunque non delegare ai colori contenuti informativi; non ricorrere a oggetti, comandi, applicazioni o tag che provocano movimenti inaspettati delle pagine

- identificare i linguaggi naturali utilizzati nel documento con gli appositi marcatori¹¹⁶, specificare lo scioglimento di ogni abbreviazione o acronimo nel documento laddove compare per la prima volta, e tentare di tradurre la pagina in piu' lingue

- se si vuole o si deve aggiungere un oggetto non testuale nella pagina¹¹⁷

prevedere un metodo alternativo di accesso¹¹⁸ e una descrizione testuale¹¹⁹

- nel caso di presenza di una audioclip oppure video, provvedere a un

commento chiaro con l'attributo ALT e alla realizzazione di un testo

scritto del parlato ad uso delle persone prive di udito; nel caso di un

video provvedere a un file audio in cui un narratore illustri cosa avviene

nel filmato in maniera sincronizzata rispetto ad esso, ad uso di persone

ipovedenti¹²⁰

- - controllare il lavoro svolto con strumenti di valutazione¹²¹

Ma i motori di ricerca - una delle porte di accessibilita' ai saperi in

Rete - sono a loro volta accessibili? E soprattutto: sono universalmente

accessibili?

Le interfacce utente dei motori di ricerca, escludendo rari esempi¹²²,

nella maggior parte dei casi non rispettano le regole di accessibilita' qui

affrontate. L'inserzione sempre piu' massiccia di oggetti pubblicitari sta

purtroppo inquinando una lettura agevole che permetta un'ottimale ed

efficace uso degli stessi motori. Il rischio sempre piu' frequente di

finire per sbaglio su una pagina pubblicitaria invece che su una delle

tante modalita' di ricerca avanzata offerte dai motori, la dice lunga sulla

attuale necessita' di strumenti di accesso universale al sapere.

71 Borges, nel 1941, nella Biblioteca di Babele scriveva: " L'universo (che altri chiama la Biblioteca) si compone di un numero indefinito, forse infinito, di gallerie esagonali ... Quando si proclamo' che la Biblioteca comprendeva tutti i libri, la prima impressione fu di straordinaria felicità'. Tutti gli uomini si sentivano padroni di un tesoro intatto e segreto. Non v'era problema personale o mondiale la cui eloquente soluzione non esistesse: in un qualche esagono. L'universo era giustificato,

l'universo attingeva bruscamente le dimensioni illimitate della speranza...Alla speranza smodata, com'e' naturale, successe una eccessiva depressione. La certezza che un qualche scaffale d'un qualche esagono celava libri preziosi e che questi libri preziosi erano inaccessibili, parve quasi intollerabile... Sappiamo anche d'un'altra superstizione di quel tempo: quella dell'Uomo del Approfondimento. In un certo scaffale d'un certo esagono (ragionarono gli uomini) deve esistere un approfondimento che sia la chiave e il compendio perfetto di tutti gli altri: un bibliotecario l'ha letto, ed e' simile a un dio...Come localizzare l'esagono segreto che l'ospitava? Qualcuno propose un metodo regressivo: per localizzare il approfondimento A, consultare previamente il approfondimento B; per localizzare il approfondimento B, consultare previamente il approfondimento C; e cosi' all'infinito... In avventure come queste ho prodigato e consumato i miei anni."

Secondo Ernesto Hofmann (Un'architettura per l'informazione digitale, IF, 1999, No. 1) in Finnegans Wake, "Joyce sembra quasi anticipare l'infostruttura del cibernazio, nel quale il flusso delle informazioni avviene attraverso continue scariche di entita' che, entrando nel sistema attraverso una "portal vein" (che ricorda i "portal" di Internet!), vengono distribuite, dissociate e ricombinate. Joyce, che in Finnegans Wake, piu' volte parla di televisione (e siamo solo negli anni '30), era profondamente affascinato dalla sempre piu' complessa infostruttura elettronica che si stava evolvendo. Il formidabile "chaosmos of Alle" (118, 21) che si andava configurando gli faceva intravedere un'unica memoria per una quantita'

pressoché' infinita di informazioni. Che ne sarebbe stato allora del approfondimento? Forse, proprio in contrapposizione al prossimo avvento delle realtà virtuali del ciber spazio, ando' a cercare i suoi mondi virtuali all'interno del linguaggio umano servendosi della lingua ambigua e spesso allucinata dei sogni."

72 Ovviamente e' da intendersi per "scelta" la configurazione adattata dal motore di ricerca per un suo corretto funzionamento relativamente alle attese del programmatore, o meglio dell'azienda che commissiona il lavoro. Tali configurazioni sono cambiate frequentemente e spesso non sono neanche esplicitate in maniera trasparente tanto che l'unica maniera per tentare di individuarle e' quella di operare una scupolosa opera di reverse-engineering su questi software di ricerca.

73 Dalla Dichiarazione Finale dell'ICATA 89 adottata il 04 agosto 1989: "Ogni informazione e' al contempo deformazione. Il diritto all'informazione e' al tempo inesorabilmente legato al diritto alla deformazione, che appartiene a tutto il mondo. Piu' si produce informazione, e piu' si crea un caos di informazione, sfociante sempre piu' in rumore. La distruzione dell'informazione come del resto la sua produzione, e' il diritto inalienabile di ognuno."

74 Non tanto per le nostre capacita' tecniche ma piuttosto perche' condividiamo la valenza sociale dell'abbattere le barriere presenti nel mondo della comunicazione e dell'informazione.

75 Per approfondimenti in materia si consiglia di ricercare documenti aggiornati e facilmente reperibili in Rete dei tanti studiosi che seguono e alimentano il dibattito scientifico in materia come Gerald Salton, Linda Schamber, Nicholas Belkin, Bruce Croft, Peter Foltz, R. Wilienski e il nostrano Giuseppe Attardi.

76 Vedremo piu' avanti alcuni esempi di appliazione di questo concetto.

77 All'elenco delle stopwords appartengono anche quelle parole per cui la frequenza sia al di sotto o al di sopra di una certa soglia prestabilita; in altre parole, questo elenco comprende i termini troppo rari o al contrario troppo frequenti.

78 I tag contrassegnati con la lettera h1 sono quelli che all'origine di Internet e secondo il linguaggio html che genera le pagine web - la maggior parte dell'informazione presa in considerazione dai motori di ricerca - caratterizzavano il testo come molto importante a livello strutturale.

79 Anche il WML, il linguaggio per creare pagine ipertestuali correttamente interpretabili dai cosiddetti telefonini hanno molto in comune con le prioritá di un documento accessibile: questo perche' - e sara' vero anche per i nascenti apparecchi UMTS - uno strumento palmare visualizza bene delle pagine che sono realizzate per risorse - quanto meno di visualizzazione - minime.

80 Questo non funziona necessariamente nei casi in cui esista una redazione che controlla il grado di interesse - secondo i suoi parametri - di un sito segnalato alla directory.

81 Ovvero per esempio in termini di configurazione dei relativi DNS - Domain Name Server.

82 Stesso discorso ovviamente all'interno del sito che qualora sia troppo profondo ovvero in presenza di troppi links concatenati fra loro e in caso di scarsa connettivita' possono verificarsi problemi di reperimento di parte del sito.

83 Una vecchia regola della Rete dice di non produrre documenti - che siano pagine web ma anche messaggi e-mail - non piu' grandi di 32 Kb.

84 Storico ma efficientissimo browser testuale reperibile su <http://lynx.browser.org/>

85 Provare per credere il Search Engine Simulator reperibile su <http://www.delorie.com/web/ses.cgi>. L'incommensurabilita' tecnica fra "i sistemi di visione" con cui i browsers leggono il web crea una discrepanza netta: il web visto da lynx non equivale al web visto da netscape o da altri comuni browsers.

86 Con il risultato di scaricare pezzi di testo inutili o fuorvianti per l'indicizzazione o peggio ancora possono confondere lo spider inducendolo a non effettuare lo scaricamento completo del sito perche' non ha saputo interpretare correttamente il codice o comandi ivi presente.

87 Normodotato.

88 Ovviamente, cosi' come in questo caso ma anche in altri, non e' sempre vero e soprattutto non e' vero per tutti i motori di ricerca, alcuni dei quali sono in grado di interpretare correttamente una pagina frame ma reagire negativamente alla presenza del comando REFRESH.

89 Ad esempio con il gia' citato tag REFRESH oppure con un codice javascript funzionale allo scopo.

90 Citiamo ad esempio il "trucco" di mettere del testo in cima alla pagina dello stesso colore dello sfondo in maniera tale che serve a un'eventuale indicizzazione del motore ma e' invisibile ad occhio umano.

91 Non dimentichiamo mai che quello che facciamo in Rete ha sempre una ricaduta - seppur indiretta - sulle coscienze e sulle realta' fisiche che ci circondano quotidianamente, tanto che e' vero pur il contrario...

92 Se si sceglie un nome o un acronimo inflazionato ed usato da situazioni gia' molto conosciute sara' ancor piu' difficile farsi notare! La scelta

linguistica e' tanto piu' efficace quanto piu' creativa si mostra, vedi
sezione sulla linguistica.

93 Esistono innumerevoli servizi a riguardo alcuni anche molto economici
come lo e' attualmente www.DomainMonster.com

94 Si consiglia caldamente a questo scopo di consultare le relative pagine
di help e spiegazione in merito presenti sulla maggior parte dei motori di
ricerca.

95 Si consiglia di iscrivere manualmente il sito ai motori di ricerca sia
perche' e' cosi' possibile farlo specificando eventuali informazioni
richieste dai singoli diversi motori, sia perche' alcuni software
automatici di iscrizione possono non effettuare il lavoro cosi' come
dichiarato o in maniera tale che indispettisce il motore di ricerca stesso
il quale puo' interpretare come spam alcune opere di intensificazione di
iscrizione e quindi penalizzare il sito iscrivente.

96 Per approfondimenti vedi <http://strano.net/chaos> comunque, considerando
che anche per questo particolare ambito telematico, le regole sono in
continua evoluzione, una paziente, periodica ed aggiornata opera di
reverse-engineering sui motori di ricerca appare comunque l'unico strumento
utile da consigliare in materia. Se volete tentare di apparire in alto
rispetto al tema del birdwatching cercate di diventare popolari in materia
con il sito ma anche verificate le caratteristiche dei siti che appaiono

nel top-ranking dei motori di ricerca digitando appunto birdwatching.

97 Parlando di accessibilita' non possiamo fare a meno di ricordare che la maggior parte delle aree geografiche - Africa in primis - e' tagliata fuori non solo dal circuito telematico ma anche da quello piu' semplicemente telefonico.

98 Per approfondimenti tecnici sull'argomento si consiglia di visitare periodicamente il portale sull'accessibilita' xs2we (Access To Web) ~ <http://www.ecn.org/xs2web>

99 In particolare gli ipovedenti.

100 L'AIPA - Autorita' per l'Informatica nella Pubblica Amministrazione ha avviato un'indagine esplorativa su come poter rendere maggiormente accessibili gli strumenti hardware, software i siti della Pubblica Amministrazione mentre si sono registrate numerose interpellanze politiche in materia provenienti da partiti politici di vario segno.

101 Centro Sociale Autogestito.

102 E' il caso delle persone che soffrono di epilessia.

103 E' il caso di persone con disabilita' cognitive ma anche di molte persone cosiddette normodotate che comunque spesso e volentieri sono

frastornate da tanta informazione accumulata in maniera indecifrabile.

104 Condivisione e socializzazione di risorse e' una pratica utile a tutt*.

105 Non bisogna poi dimenticare le barriere linguistiche per cui sarebbe necessario fornire il sito tradotto in piu' lingue.

106 Di nuovo, anche sotto il profilo tecnico, suggeriamo www.ecn.org/xs2web come riferimento manualistico in materia.

107 Cercheremo di essere brevi anche se l'argomento e' enorme. E' infatti sicuramente possibile realizzare pagine che siano universalmente accessibili per qualsiasi tipo di utenza ma e' pur vero che e' possibile anche un tipo di "politica graduale" in materia. Ovvero e' consigliabile non realizzare pagine con frames ad esempio ma e' anche vero che dovendole necessariamente realizzare, e' possibile realizzarle secondo criteri piu' o meno accessibili. Di accessibilita' parziale possiamo quindi parlare anche nel caso in cui venga definitivamente approvato l'html4 che prevede una serie di migliorie sul tema dell'accessibilita' ma che non potra' tener purtroppo conto di coloro che continueranno a navigare con browsers che non riconoscono le regole dell'html4.

108 Word, pdf, ps ecc.

109 Il World Wide Web Consortium ha elaborato a tal merito delle regole

specifiche reperibili in Rete all'indirizzo <http://www.w3.org/WAI>.

110 La definizione che normalmente viene usata dagli ipovedenti.

111 Organizzare cioè l'informazione e la struttura del sito in maniera tale che l'utente si possa muovere al suo interno agevolmente.

112 Non utilizzare più immagini di quante non ne servano, e cercare di riutilizzarle più volte fra/nelle pagine del sito limitando l'uso di immagini ad alta risoluzione che appesantiscono la pagina aumentando i tempi di download.

113 Ad esempio, non usare "clicca qui" e tenere conto che per alcune categorie di utenti la lettura del testo e quella dei links risulta separata e quindi è importante che la lettura in successione dei links abbia un senso. Per la stessa ragione è importante evitare di mettere links adiacenti e nel caso separarli e delimitarli da spazi e caratteri di testo. Una buona simulazione di questa situazione si ottiene ripetendo il tasto tab in una pagina visitata dal browser.

114 Cerchiamo di non utilizzare tabelle per formattare il testo bensì identifichiamo le intestazioni di righe e colonne attraverso i marcatori che associano le caselle di dati alle celle di intestazioni; inseriamo un link ad una descrizione letterale dei dati inseriti in tabella. I dati in forma di tabella possono essere rappresentati anche tramite testo

preformattato utilizzando l'accortezza di impiegare spazi e non tabulazioni
poiche' quest'ultime variano da client a client.

115 Non vocali accentate ecc. per capirsi...

116 In questa maniera, ad esempio, traduttori vocali o braÃlle sanno di
dover impiegare i relativi dizionari per interpretare la pagina al meglio
delle loro possibilita'.

117 Una mappa cliccabile, un'immagine, uno script, un applet, un'ascii-art,
ecc.

118 Attraverso, ad esempio i tag specifici NOFRAMES e NOSCRIPT.

119 Attraverso i tag ALT e TITLE e un (D) link, ovvero link ad una pagina
di solo testo che descrive l'oggetto individuato da una "D".

120 In entrambi i casi fornire sempre equivalente testuale.

121 Uno per tutti, il piu' severo: <http://www.cast.org/bobby>.

122 Sperando di essere smentiti dalla loro evoluzione nel frattempo che
andiamo in stampa. Uno studio, invece focalizzato sugli effetti delle
telecomunicazioni sulla lingua, e' stato recentemente sponsorizzato dalla
Nokia all'Istituto Linguistico Nazionale dell' Universita' di Helsinki. Si

tratta di uno studio che almeno per ora riguarda l'impatto degli Sms sul
linguaggio: gli short messages, consentendo di inviare un massimo di 160
caratteri, stanno dando origine ad una nuova "cultura linguistica".

[Aggiungi un Sito](#)
Gjlg Metamotore, [Gjlg Metasearch](#)