

Appunti di Web Design

Giacomo Zanatta

April 9, 2018

Contents

1	Introduzione	4
1.1	Evoluzione dei linguaggi	4
1.2	Siti Responsive	4
2	Graphics VS Web	4
2.1	World Wide Web	4
2.1.1	Caratteristiche di successo	4
2.1.2	Evoluzione del Web	5
2.1.3	Epoche del web	5
2.2	Desing pre-web (editoria)	5
2.2.1	Dogmi del graphic design	5
2.3	Progettare per l'ignoto	5
2.4	Il carattere digitale	6
2.4.1	Rasterizzazione	6
2.4.2	Unità di misura del carattere	6
2.4.3	Se l'autore non specifica il font	7
2.4.4	Testo come immagine	7
2.4.5	Il futuro del font	7
2.5	Colore	8
2.5.1	Modelli di colore	8
2.5.2	Spazio colore	11
2.5.3	Pixel	12
2.5.4	Colore RGB digitale	12
2.5.5	Discretizzare il reale	12
2.5.6	Colori per il Web	12
2.5.7	Gamma	13
2.6	Progettare il layout	13
2.6.1	Prodotto editoriale stampato	13
2.6.2	Vantaggi della gabbia tipografica	13
2.6.3	Progettare layout web	14
2.6.4	Area sicura	14
2.6.5	Layout flessibili e fissi	14

3	Pianificare un sito web	15
4	Architettura dell'informazione	15
4.1	I 3 cerchi dell'architettura dell'informazione	16
4.2	Bisogni e comportamenti degli utenti	16
4.2.1	Che cosa vogliono gli utenti?	17
4.2.2	Comportamenti dell'utente	17
4.2.3	Modello berry-picking	17
4.2.4	Modello pear-growing	17
4.3	Progettare la struttura dell'informazione	18
4.3.1	Problemi nell'organizzazione dell'informazione	18
4.4	Organizzare l'informazione	18
4.4.1	Schemi organizzativi	18
4.4.2	Strutture organizzative	19
4.4.3	Approccio top-down	19
4.4.4	Approccio bottom-up	20
4.4.5	Creare sistemi organizzativi coesi	21
5	Sistemi di navigazione	21
5.1	Il design di sistemi di navigazione	21
5.2	Costruire il contesto	21
5.3	Navigation Stress Test	22
5.4	Tipologie di Sistemi di Browsing	22
5.4.1	Sistema Gerarchico	22
5.4.2	Sistema Globale	22
5.4.3	Sistemi Locali	22
5.4.4	Sistemi Ad hoc - Trasversali - Contestuali	22
5.5	Elementi di browsing integrati	23
5.5.1	Menu di navigazione	23
5.5.2	Shortcuts	23
5.5.3	Menu pop-up e pull-down	23
5.5.4	Drawer	23
5.6	Sistemi di browsing	23
5.7	Elementi di browsing remoti	24
5.7.1	Indice dei contenuti	24
5.7.2	Indice analitico	24
5.7.3	Mappa del sito	24
5.7.4	Tour guidato	24
5.8	Sistemi di browsing personalizzati	24
5.9	Sistemi di navigazione sociale	25
5.9.1	Tag clouds	25
5.9.2	Interfaccia per l'inserimento di tag	25
5.9.3	Integrazione nel layout delle interfacce di navigazione sociale	25
5.9.4	Utenti, risorse e annotazioni	25
5.9.5	Fake tagcloud	25

6	Sistemi di ricerca	26
6.1	Quando implementare un sistema di ricerca?	26
6.2	Zone di ricerca	26
6.3	Pagine di destinazione e navigazione	26
6.4	Componenti di una pagina	26
6.5	Presentare i risultati	26
6.5.1	Quali?	26
6.5.2	Quanti?	27
6.5.3	Elencare i risultati	27
6.5.4	Raggruppare i risultati	27
6.5.5	Esportare i risultati	27
6.6	L'interfaccia di ricerca	28
6.7	L'interfaccia per i risultati	28
6.8	Interfacce di ricerca basate sulla classificazione sociale	28
6.9	Trovabilità	28
7	HTML	28
7.1	Timeline di HTML	29
7.2	Tag HTML	29
7.2.1	Elementi vuoti	29
7.2.2	Nidificazione di Tag	30
7.2.3	Attributi di tag	30
7.3	Caratteri ignorati	30
7.4	Struttura di un documento XHTML 1	30
7.4.1	Prologo XML	30
7.5	Specifica dei caratteri	31
7.6	DTD (Document Type Definition)	31
7.7	Il tag <i><html></i>	31
7.8	Conformità	32
7.9	Il tag <i><head></i>	32
7.9.1	Il tag <i><meta></i>	32
7.10	Il tag <i><body></i>	33
7.11	HTML 5	33
7.11.1	Struttura di un documento HTML 5	33
7.11.2	Doctype	33
7.11.3	Codifica dei tag e attributi	33
7.11.4	Obsolescenza	34

1 Introduzione

1.1 Evoluzione dei linguaggi

- Layout solo HTML
- Layout ibridi (tabelle + regole css per la presentazione)
- Layout CSS (div, span + css evoluto per caratteristiche posizionali di presentazione)
- Layout CSS + media queries (responsive)

Separare contenuto e presentazione significa poter cambiare la visualizzazione del sito senza doverne cambiare il contenuto

1.2 Siti Responsive

- Permettono una presentazione flessibile, che si adatta alla tipologia di dispositivo utilizzato.
- Senza un sito responsive, possono sorgersi problemi se l'accesso ad un sito avviene attraverso l'uso di una rete sociale (esempio, se metto un link di un sito su facebook, se il sito non è responsive posso aver problemi a visualizzarlo correttamente se sono da mobile)
- Per ottenere un sito responsive posso utilizzare diverse tecniche: posso usare un framework, un CMS, oppure usare direttamente HTML, CSS e JavaScript. Inoltre posso usare tecniche di riconoscimento del browser per indirizzare l'utente verso la versione del sito adatta per il suo dispositivo o fornire una versione unica per i contenuti e forme di presentazione differenziate, opportunamente selezionate dai browser.

2 Graphics VS Web

2.1 World Wide Web

Il web nasce nel 1993, per permettere di condividere rapidamente in modo centralizzato i risultati scientifici di gruppi di lavoro. Il web permette l'ipertestualità e l'ipermedialità. Proprio su queste caratteristiche sono nati applicativi commerciali (come Hypercard e Toolbox) e sono stati fatti studi teorici (Nelson e Engelbart).

2.1.1 Caratteristiche di successo

- Paradigma intuitivo per l'utente
- Condivisione dell'informazione

- Multiplatforma
- Linguaggio di facile utilizzo e basato su marcatori (tag)

2.1.2 Evoluzione del Web

Inizialmente era una rete locale, poi è diventata globale.

Ora nel web possono accederci non solo utenti che lavorano in ambito scientifico ma qualsiasi persona, da qualsiasi parte del mondo.

Il web inoltre presenta diversi ambiti applicativi: può essere utilizzato come frontend di sistemi informativi oppure come contenitore di contenuti.

2.1.3 Epoche del web

1. 1993: Far Web
2. 1994-1995: Monolito
3. 1996 - ...: Approccio interdisciplinare

2.2 Desing pre-web (editoria)

- Output cartaceo fisso, con metodologie non informatizzate (stampe d'autore...)
- Output cartaceo fisso, con metodologie informatizzate (giornali, riviste, libri...)
- Output elettronico fisso, con metodologie informatizzate (cd-rom, pubblicazioni elettroniche...)

2.2.1 Dogmi del graphic design

- Carattere tipografico inalterabile (font PostScript)
- Colore inalterabile
- Inalterabilità della composizione visuale

2.3 Progettare per l'ignoto

Per progettare per il web è necessario tenere conto di molti fattori:

- Versione del browser
- Piattaforme HW
- Preferenze utente
- Velocità connessione
- Caratteri tipografici

- Colori
- Dimensione della finestra di visualizzazione e layout

Per sopravvivere all'ignoto è necessario abbandonare la pretesa del controllo assoluto. Disegniamo quindi strutture e imponiamo set di regole, ma non dobbiamo raggiungere il controllo assoluto al singolo pixel.

2.4 Il carattere digitale

La rappresentazione digitale del carattere è soggetta a molte limitazioni. Viene rappresentato attraverso un insieme di pixel che fanno riferimento ad una griglia di base. Nei sistemi operativi più vecchi, i caratteri tipografici vengono rappresentati come mappe di punti (font bitmap) pre-disegnate a dimensioni specifiche.

Nei sistemi moderni vengono utilizzati font outline, ossia rappresentati come primitive matematiche.

2.4.1 Rasterizzazione

Il processo di rasterizzazione di un font consiste nel convertire il testo da una descrizione vettoriale (font outline) ad una descrizione bitmap o raster.

Spesso si usano tecniche di antialiasing sul testo che deve essere letto sullo schermo, per renderlo più gradevole e leggibile.

È possibile usare anche l'hinting, che rende il font più gradevole e leggibile per una particolare dimensione usando informazioni pre-calcolate. Recentemente viene usato il subpixel rendering, nel quale si utilizzano le tre componenti RGB per aumentare la risoluzione dell'immagine.

2.4.2 Unità di misura del carattere

Il carattere tipografico può essere specificato attraverso diversi sistemi.

In ogni caso la presentazione del carattere su schermo passa attraverso il processo di rasterizzazione, che si basa sulla specifica dell'unità di misura, ma anche su alcune assunzioni. Le unità di misura utilizzate sono:

- Pixel (px): unità minima di colore visualizzabile su schermo.
- Point (pt): misura tipografica tradizionale (72 punti per pollice).
- Pica (pc): 1 pica sono 12 punti (1/6 di pollice).
- Em (em): unità relativa che corrisponde alla larghezza della lettera 'M' nel carattere utilizzato.
- Ex (ex): si basa sull'altezza della lettera 'x' nel carattere utilizzato (circa metà em)

- Inches (in): unità di misura standard negli USA.
- Millimetri (mm)
- Centimetri (cm)

2.4.3 Se l'autore non specifica il font

In assenza di specifiche, la selezione del tipo e della dimensione dei font della pagina è a carico del browser. In questo caso, possono sorgere risultati diversi a seconda del browser o del S.O. utilizzato. Negli ultimi anni, comunque, si sta cercando di attuare un processo di standardizzazione.

Vengono utilizzati due font:

1. Font proporzionale: viene allocata una diversa quantità di spazio orizzontale per ogni carattere, sono più facili da leggere (Times, Helvetica, Arial).
2. Font a larghezza fissa: viene allocata la stessa quantità di spazio orizzontale per tutti i caratteri. È utile per incolonnare i caratteri, ad esempio per mostrare sulla pagina web frammenti di codice (Courier, Monaco).

Ovviamente (per quasi tutti i browser) è possibile cambiare il font di default dalle impostazioni.

2.4.4 Testo come immagine

È possibile codificare il testo come una immagine. Questo porta dei vantaggi: abbiamo il controllo assoluto (sappiamo che la visualizzazione sarà esattamente come la vogliamo, su qualsiasi browser) ma anche dei svantaggi (le immagini sono lente a caricarsi sul browser, su browser non grafici l'immagine non sarà visualizzata e il testo non può essere indicizzato).

2.4.5 Il futuro del font

Da un po' di tempo sono disponibili i web fonts, ossia font direttamente accessibili da un server online che ne permette l'utilizzo e il download.

In questo modo abbiamo la possibilità di scaricare un'ampia famiglia di font vettoriali attraverso un link da una pagina web. Al font possiamo applicare effetti, come ad esempio l'ombreggiatura.

Google fornisce font gratuiti, ma esistono online alcuni store che permettono di scaricare font a pagamento e su licenza.

Un problema da tenere in considerazione sono le performance: è necessario tenere conto della velocità di caricamento del font e inoltre è necessario implementare una soluzione di backup per i browser che non riescono a visualizzare questi font.

2.5 Colore

Il colore inoltre è in funzione di: risorse HW di sistema, sistema operativo e browser (il quale possiede risorse autonome per effettuare il rendering del colore su sistemi con limitate capacità grafiche).

2.5.1 Modelli di colore

Un modello di colore è un modello matematico astratto che permette di rappresentare i colori in forma numerica. Esistono diversi modelli di colore, tra cui:

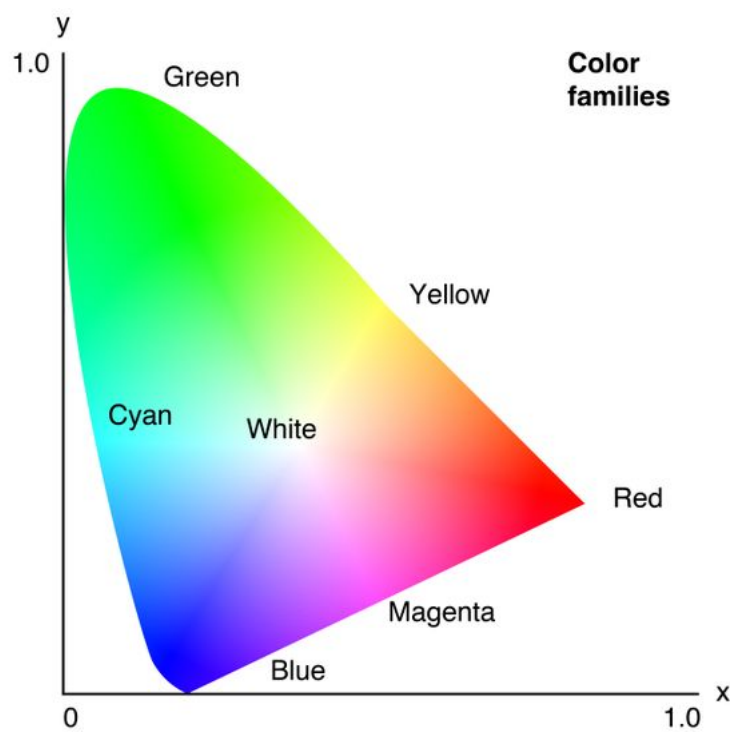
1. **CIE Yxy** : è un sistema che simula bene il processo visivo, e per definire un colore viene utilizzato un triangolo.

Questo triangolo descrive lo spazio colore tramite due variabili cromatiche x e y .

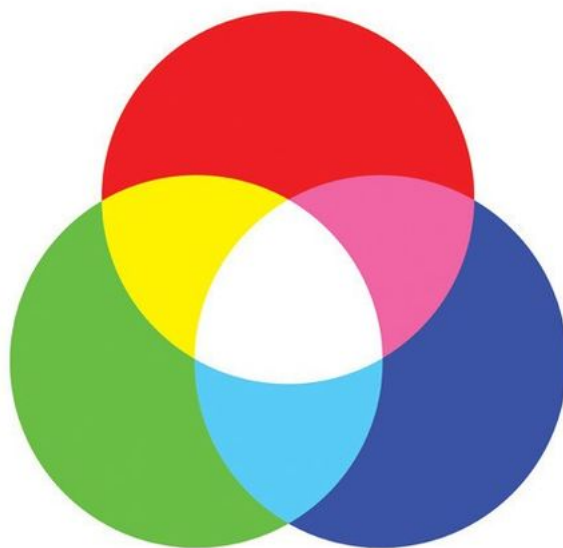
Sul piano cartesiano è situata una curva a ferro di cavallo, sul cui bordo sono situati i colori puri identificati dalle lunghezze d'onda. Più ci si sposta verso il centro del grafico, più la saturazione si riduce e il colore diventa sempre più neutro.

Il valore x indica l'importanza della componente rossa del colore nei confronti delle componenti verde e blu, ed è inferiore a 1. Il valore y invece indica l'importanza della componente verde nei confronti delle componenti rossa e blu, anche questa inferiore a 1.

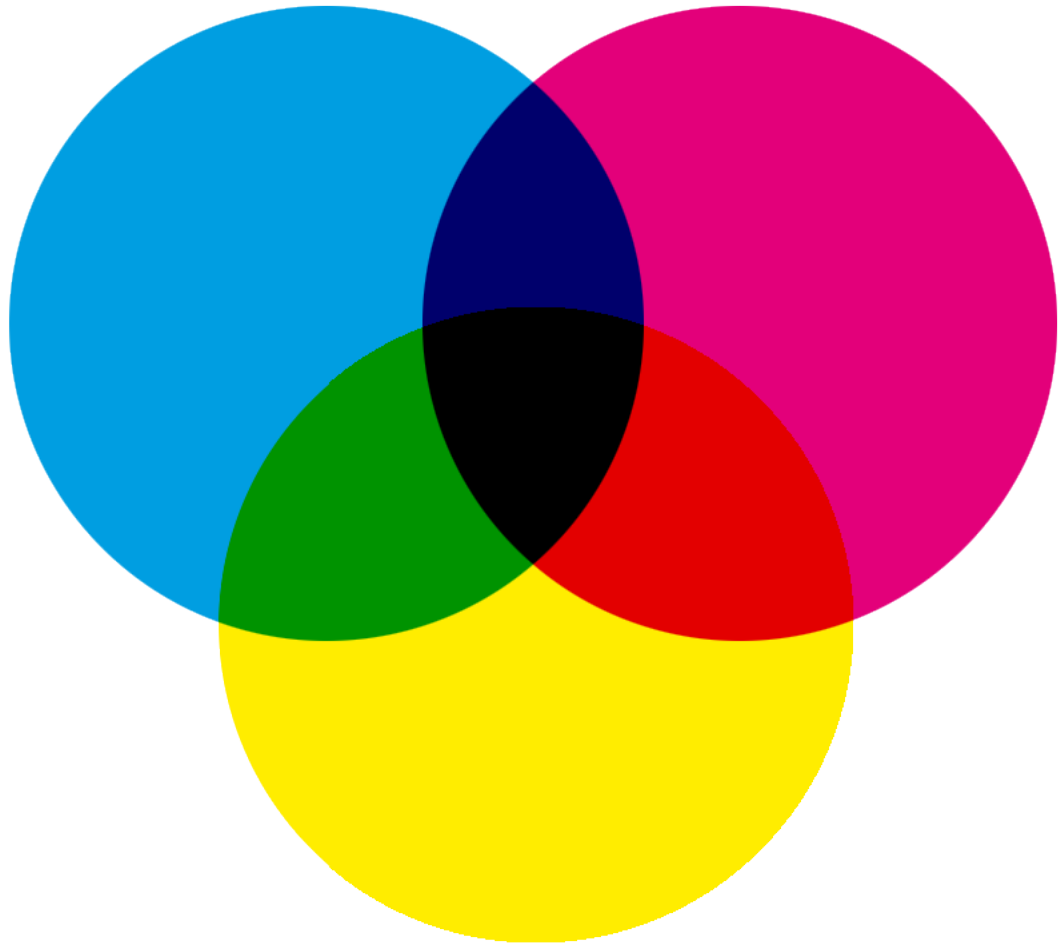
Y (maiuscolo) indica la luminosità (compresa tra 0 e 100).



2. **RGB**: una vasta percentuale dello spettro visibile può essere rappresentata miscelando i 3 componenti della luce colorate Rosso, Verde e Blu in diverse proporzioni e intensità.
Vengono chiamati colori addittivi perchè se sommiamo questi 3 colori generiamo il colore bianco. I colori addittivi vengono usati per l'illuminazione, i video e i monitor.



3. **CMYK**: questo modello si basa sulla capacità di assorbimento della luce dell'inchiostro sulla carta. Quando la luce bianca colpisce gli inchiostri traslucidi, una parte dello spettro viene assorbita e una parte viene riflessa all'occhio.
I pigmenti puri di Cyan, Magenta e Giallo si sommano per assorbire tutto il colore e produrre il Nero: per questo sono chiamati colori sottrattivi. La combinazione di questi inchiostri per riprodurre il colore viene chiamata stampa in quadricomia.



2.5.2 Spazio colore

Lo spazio colore indica l'intervallo dei colori che possono essere visualizzati o stampati.

Lo spettro dei colori percepiti dall'occhio umano è maggiore di qualsiasi metodo di riproduzione del colore.

Fra i vari modelli del colore, CIE ha lo spazio colore più ampio, e comprende gli spazi di colore di RGB E CMYK.

Ci sono diversi spazi colore RGB: ad esempio, sRGB (disegnato in relazione alle possibilità di visualizzazione dei monitor CRT, gamma molto ristretta), Adobe RGB (che include la maggior parte dei colori ottenibili sulle stampanti CMYK ma utilizzando colori primari RGB, gamma molto più ampia di sRGB), Adobe WWide Gamut RGB (versione estesa di Adobe RGB).

È necessario tenere in considerazione che questi sono modelli teorici rispetto ai quali i singoli dispositivi riescono ad adeguarsi in modo diverso.

2.5.3 Pixel

Per controllare il colore di ogni pixel sullo schermo il sistema operativo deve dedicare una piccola quantità di memoria ad ognuno di essi.

La memoria dedicata allo schermo è una memoria distinta, che risiede nella scheda grafica e ci si riferisce con Video RAM (VRAM).

2.5.4 Colore RGB digitale

Per ogni pixel, abbiamo:

- HIGH COLOR: 16 bit, 5 bit per componente (il sedicesimo bit utilizzato per altri scopi).
- TRUE COLOR: 24 bit, 8 bit per ogni componente. Se si utilizzano 32 bit per pixel, i bit aggiuntivi vengono riservati per velocizzare le operazioni della scheda grafica o per informazioni di mascheratura/trasparenza.
- DEEP COLOR: 16/32/48/64 bit per componente.

Colore RGB a 24 bit

Con schede grafiche RGB a 24 bit, viene assegnato un valore di intensità a ogni pixel compreso tra 0 (nero) a 255 (bianco) per ognuna delle componenti RGB di un'immagine a colori. Le immagini RGB utilizzano 3 colori per riprodurre fino a 16,7 milioni di colori sullo schermo.

2.5.5 Discretizzare il reale

La realtà percepita dall'occhio umano è a tono continuo. Catturare un'immagine con strumenti digitali vuol dire discretizzare l'informazione contenuta nella scena reale.

2.5.6 Colori per il Web

Per specificare i colori RGB è quello di utilizzare il valore numerico della tripletta di componenti, convertiti in notazione esadecimale.

Su sistema hardware dotati capacità grafiche inferiori (8-16 bit) i colori provenienti da uno spazio colore true space (ossia a 24 bit) vengono approssimati. Per operazioni a livello di sistema i computer utilizzano un set specifico di 256 colori, chiamato palette di sistema.

Ogni sistema operativo ha la sua palette. Per questo motivo è stata introdotta una web-safe palette, che consiste di 216 colori comuni alle palette di sistema Win e Mac.

Tutti i colori della web palette sono combinazioni dei valori esadecimali 00,33,66,99,CC,FF. L'utilizzo di colori che non appartengono alla web palette su computer con scarse risorse grafiche può portare ad una approssimazione del colore (dithering) e a risultati sgradevoli.

2.5.7 Gamma

La gamma denota la luminosità complessiva del display. Più è alto il valore, meno luminosa è l'immagine sul display. Notare che ogni piattaforma ha una gamma diversa, quindi immagini create su Mac appaiono più scure sui sistemi Win.

2.6 Progettare il layout

Nel design tradizionale, le dimensioni del supporto vengono stabilite a priori e costituiscono un vincolo immutabile durante la progettazione del layout.

Nel design tradizionale viene definita la gabbia tipografica, che partiziona lo spazio disponibile in aree omogenee le quali conterranno testo, grafica, o (nel caso di supporti elettronici) elementi multimediali. È un elemento caratterizzante del design complessivo, quindi da mantenere costante durante la pubblicazione.

2.6.1 Prodotto editoriale stampato

Nei giornali è presente un modulo verticale (principalmente su 8 colonne) che costituisce la base del layout del quotidiano.

Tutti gli altri elementi sono costruiti basandosi su questo modulo di base, rispettando allineamenti e regole di simmetria.

Per convenzione, la pubblicità viene posizionata nei due riquadri superiori e nella base del layout. L'articolo di fondo è sempre posizionato a sinistra. La zona superiore viene riservata alle notizie generali rilevanti mentre quella inferiore alla cronaca locale.

2.6.2 Vantaggi della gabbia tipografica

- Aspetti funzionali: maggiore facilità nel reperimento dell'informazione da parte del lettore e maggiore facilità nell'interazione con i prodotti editoriali interattivi da parte dell'utente.

- Aspetti comunicativi: è l'elemento fondamentale dell'identità di un progetto grafico.
- Creatività: è possibile coniugare rigore e creatività, introducendo eccezioni rispetto alla regola data.

2.6.3 Progettare layout web

- Supporto: problematico assicurare la corrispondenza tra la dimensione complessiva della pagina e l'area visibile all'utente.
- Controllo del layout: nelle prime versioni di HTML non era possibile alcuna forma di controllo del layout della pagina. Il CSS permette un controllo accurato del layout, permettendo anche di ottenere layout diversi per media diversi.

2.6.4 Area sicura

- Per la visualizzazione: numero di punti dello schermo (pixel) disponibili per visualizzare l'informazione di una pagina web.
- Per la stampa: numero di punti dello schermo stampabile su carta.

L'area sicura dipende da configurazione HW e SW, browser e preferenze dell'utente. Nell'area sicura è meglio mettere gli elementi informativi fondamentali, gli artefatti fondamentali per l'interazione con il sito, e gli elementi grafici caratterizzanti.

2.6.5 Layout flessibili e fissi

- Layout flessibile: dimensioni delle aree fissate usando misure relative (es percentuale) come anche le dimensioni del carattere, per permetterne uno zoom e un rimpicciolimento.
Vantaggi: pagina si adatta al display e alle preferenze dell'utente.
Svantaggi: in alcuni casi possono verificarsi righe troppo lunghe e non facilmente leggibili.
- Layout fissi: dimensioni fissate usando misure assolute (es pixel).
Vantaggi: risultato unico, maggior controllo.
Svantaggi: problemi su determinate configurazioni HW e SW (ad esempio IE fino alla versione 6 non permette all'utente di riscalare un testo le cui dimensioni siano state fissate in pixel. Non viene garantito il controllo sul carattere tipografico (come descritto nella sezione font).

La scelta della tipologia di layout dipende da diversi fattori, ad esempio gli utenti, l'ambiente e la tipologia di servizio offerto. È preferibile l'adozione di layout flessibili, che permettono un migliore adattamento a dispositivi eterogenei.

3 Pianificare un sito web

Progettare e sviluppare siti web è un'attività complessa che richiede un team di sviluppo interdisciplinare.

- Project Manager: responsabile coordinamento team, si occupa della schedulazione dei task e del controllo del budget.
- Esperto in marketing: identifica obiettivi del sito e utenti.
- Information Designer: è il responsabile del sistema di strutturazione, classificazione, ricerca e navigazione all'interno del sito.
- Informatico: amministra il server, sviluppa e tiene aggiornati servizi e app.
- Web Designer: progettista del design e del layout. Crea relazioni tra gli elementi del sito.
- Giornalista/Responsabile editoriale: prepara e adatta i testi da inserire nel sito.
- Esperto in usabilità: responsabile valutazione usabilità dei prototipi e del sito finale.

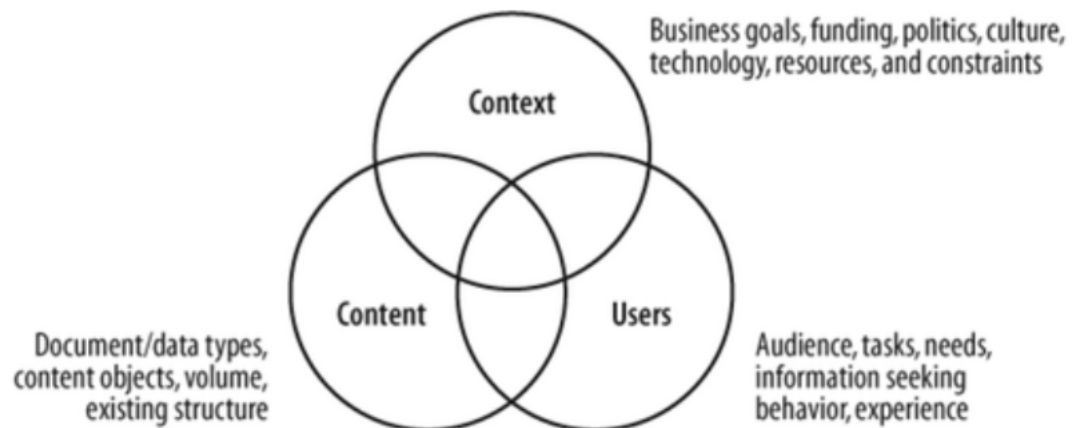
4 Architettura dell'informazione

Cos'è l'architettura dell'informazione?

- il design strutturale di ambienti informativi condivisi
- la combinazione di organizzazione, etichettatura, ricerca e sistemi di navigazione relativi a siti web e intranet
- l'arte e la scienza di dare forma a prodotti ed esperienze informative per supportare l'usabilità e la trovabilità.

Nell'architettura dell'informazione troviamo Sistemi (sistemi di ricerca, di navigazione, reti semantiche) e prodotti (strutture organizzative, vocabolari controllati, metadati, layout).

4.1 I 3 cerchi dell'architettura dell'informazione



- **Contenuti:** includono documenti, applicazioni, servizi, schemi e metadati. I parametri da considerare includono il produttore e il proprietario dei contenuti, il formato, la granularità, i metadati, il volume e la dinamicità dei contenuti.
- **Contesto:** i siti vengono definiti all'interno di un determinato contesto, con una missione specifica, obiettivi, strategia, staff, processi e procedure. L'architettura dell'informazione di un sito deve fornire un'immagine tangibile dell'organizzazione che lo promuove.
- **Utenti:** esistono diversità nelle preferenze degli utenti e nei comportamenti.

Contenuto, utenti e contesto sono componenti interdipendenti nell'ambito del processo di definizione dell'informazione.

4.2 Bisogni e comportamenti degli utenti

Modello informativo too-simple

- utente pone una domanda
- accade qualcosa
- utente riceve risposta

- fine ricerca

I problemi che possono sorgere sono che l'utente non sempre sa quello che vuole. Inoltre spesso la ricerca termina con un insuccesso, o un successo parziale. Il contesto viene ignorato.

4.2.1 Che cosa vogliono gli utenti?

Ci sono diversi tipi di bisogno, possiamo definirli con la metafora della pesca:

- **Il tiro perfetto:** quando gli utenti sanno quello che stanno cercando.
- **Trappola per aragoste:** gli utenti non hanno un'idea precisa di quello che stanno cercando, si aspettano di trovare qualcosa e di imparare qualcosa dal processo esplorativo, che possa guidarli verso una nuova ricerca.
- **Pesca con la rete:** utenti vogliono esaminare ogni elemento relativo ad un particolare argomento
- **Boa di segnalazione:** gli utenti vogliono ritrovare un elemento informativo utile

4.2.2 Comportamenti dell'utente

Gli utenti trovano l'informazione mediante:

1. searching (ricerca)
2. browsing (navigazione)
3. asking (facendo domande)

Searching, browsing e asking spesso sono integrati nella stessa sessione di lavoro, oppure utilizzati in iterazione.

4.2.3 Modello berry-picking

Gli utenti partono con un bisogno informativo, formulano una query (richiesta informativa), e si muovono iterativamente attraverso percorsi potenzialmente complessi, raccogliendo progressivamente elementi informativi. Se il comportamento degli utenti rispecchia questo modello, è necessario che sia possibile spostarsi facilmente tra searching e browsing.

4.2.4 Modello pear-growing

Gli utenti partono con uno o pochi documenti che corrispondono esattamente a quello di cui hanno bisogno per cercare altri documenti con quelle caratteristiche (es. pagine simili di Google, oppure ritrovare documenti indicizzati con le stesse keyword).

4.3 Progettare la struttura dell'informazione

Internet dà la libertà di pubblicare l'organizzazione e la responsabilità di organizzarla (nel passato quest'ultimo compito era svolto da figure professionali come i bibliotecari).

4.3.1 Problemi nell'organizzazione dell'informazione

- **Ambiguità:** i sistemi di classificazione sono basati su un linguaggio (che può essere ambiguo), inoltre classificare oggetti e concetti astratti può essere difficoltoso.
- **Eterogeneità:** molti siti web sono eterogenei, forniscono cioè l'accesso a documenti a diversi livelli di granularità e formati multipli.
- **Differenze di prospettiva:** è indispensabile mettersi nei panni dell'utente.
- **Diversità di politiche**

4.4 Organizzare l'informazione

4.4.1 Schemi organizzativi

Permettono di suddividere gli oggetti informativi in raggruppamenti logici basandosi su proprietà caratterizzanti dei singoli oggetti. Sono suddivisi in schemi organizzativi esatti e ambigui.

- **Esatti:** l'informazione viene suddivisa in sezioni mutuamente esclusive. La progettazione e la manutenzione è facile, ma richiedono che l'utente conosca il nome specifico della risorsa che sta cercando (know-item searching). Degli esempi sono: schema alfabetico, schema cronologico, schema geografico.
- **Ambigui:** sono utili per realizzare e soddisfare uno stile di ricerca basato sulla serendipità. L'informazione è suddivisa in categorie nelle quali può essere difficile collocare l'oggetto. Sono più importanti e utili dei schemi organizzativi esatti, nel caso in cui non sappiamo che cosa stiamo cercando (exploratory searching).

In questo caso la serendipità può essere supportata implementando una ricerca iterativa e interattiva, e coinvolgendo meccanismi di apprendimento associativo.

Gli schemi organizzativi ambigui più comuni sono: schemi per argomento (topical, andando a definire l'universo degli argomenti che l'utente si aspetta di trovare), schemi orientati al compito (task oriented, dove il contenuto e le applicazioni sono organizzati come collezioni di processi, funzioni o compiti), schemi specifici per audience, schemi metaforici (utilizzati per far comprendere concetti nuovi collegandoli a concetti familiari, da usare con cautela ed è necessario evitare problemi di inconsistenza), schemi ibridi (mix di tutto).

Le facets (sfaccettature) permettono di accedere allo stesso informativo da angolature (schemi organizzativi) diversi. Definire più facets permette all'utente una maggiore interazione in quanto può navigare in diversi modi, è possibile inoltre associare una caratteristica dell'oggetto al facets (ad esempio un negozio di vestiti può essere navigato per marca, colore, taglia, ...).

4.4.2 Strutture organizzative

Definiscono le tipologie di relazione tra elementi o gruppi di oggetti dell'universo informativo. Esistono diverse strutture organizzative, ognuna con i suoi punti di forza e debolezze.

- **Sequenze:** l'informazione è messa in sequenza. Le sequenze lineari sono adatte a siti didattici in cui l'utente deve procedere ordinatamente attraverso un insieme di materiali.
- **Gerarchia:** c'è un livello di parentela tra le informazioni, come una struttura ad albero. Gli utenti che utilizzano questa organizzazione sono in grado di sviluppare più facilmente un modello mentale della struttura del sito.
Le categorie gerarchiche dovrebbero essere mutuamente esclusive ma non è illegale posizionare un numero limitato di oggetti informativi in più di una categoria (cross listing) per creare una struttura poligerarchica. Posizionare un numero eccessivo di elementi in più categorie porta una perdita di valore di questa struttura.
È necessario inoltre definire un equilibrio tra ampiezza (numero di opzioni ad ogni livello) e profondità (numero di livelli) della gerarchia (non oltre 10 opzioni, e 4 o 5 livelli).
- **Ipertesto:** innovativa modalità non lineare per strutturare l'informazione. Le unità informative possono essere collegate gerarchicamente, non gerarchicamente o in entrambe le modalità.
Può essere un ostacolo per la formazione di un modello mentale del sito
- **Database:** è una collezione di record, dove ogni record ha un numero n di campi associati. I vantaggi sono: ricerca per campo, metadati associati ai dati, vocabolario controllato (imponibile un grado di consistenza che può risultare utile nella ricerca e navigazione), gestione più facile dei contenuti. Gli svantaggi sono che la strutturazione a record è rigida, e può essere costoso disporre ogni elemento del sito in un database.
È utile usare un database per rappresentare una collezione di oggetti con le stesse proprietà. Molto spesso un database viene usato per modellare anche gli altri elementi strutturali del sito e non solo il contenuto principale.

4.4.3 Approccio top-down

Le strutture organizzative viste precedentemente vengono costruite usando un approccio top-down: l'information designer fornisce una soluzione relativa ad

un dominio informativo per un determinato sito web.

Spesso, però, non tutti gli utenti trovano una corrispondenza tra la struttura progettata dal designer e il proprio modello mentale relativo al sito. Per limitare questi problemi è possibile fornire soluzioni poligerarchiche, motori di ricerca interni, mappe del sito oppure permettere agli utenti di costruire la gerarchia informativa con tecniche quali il free listing e il card sorting.

- **Free listing:** permette di coinvolgere gli utenti nella definizione dei contenuti del dominio. Viene richiesto di formulare un elenco di elementi informativi a partire dalla descrizione di un tema fornita da chi gestisce il test.
- **Card sorting:** permette di coinvolgere gli utenti nella strutturazione dei contenuti. Viene richiesto di suddividere in gruppi una lista di schede etichettate. Si può usare l'open card sorting (agli utenti vengono fornite schede con etichette relative al contenuto del sito ma senza gruppi prestabiliti; l'utente deve quindi creare i gruppi e assegnare le etichette descrittive dei gruppi stessi), o il closed card sorting (agli utenti vengono fornite sia le schede con etichette dei contenuti del sito, sia i gruppi con le etichette già definite; l'utente deve riempire i gruppi con le schede). L'open card viene usato per creare un'architettura informativa, per ottenere feedback su quali contenuti vengono inseriti in uno stesso gruppo, e capire quali etichette vengono utilizzate dagli utenti per descrivere il contenuto. Il closed card sorting, invece, risulta utile per testare il design di un'architettura dell'informazione, e per ricevere feedback sull'efficacia delle etichette.

Il free listing e il card sorting mirano ad aumentare la trovabilità degli elementi di un sito web. Per validare la gerarchia di un sito viene usato il tree testing: all'utente viene proposto una serie di task consistenti nel trovare un determinato elemento informativo facente parte di una struttura gerarchica. L'utente una volta raggiunto l'elemento informativo che reputa corretto deve confermare la scelta. In un tree testing vengono considerati: tempo impiegato, precisione (percentuale di utenti che non sono tornati indietro) e percentuale di successo.

4.4.4 Approccio bottom-up

Si costruisce la struttura informativa dal basso, dando modo all'utente di marcare gli elementi informativi che sta navigando con un set di tag. Una tecnica è il free tagging, ossia categorizzazione collaborativa di elementi informativi del web che può far emergere una forma di organizzazione complementare (o alternativa) all'approccio top-down.

Si può usare una folksonomia allargata (accettati tutti i possibili tag, anche ripetuti, usata da Delicious) o una folksonomia ristretta (per ogni risorsa il sistema non accetta che un utente inserisca tag già inseriti da altri utenti, usata da Flickr). I vantaggi di questo approccio sono: l'approccio è meglio di niente, in quanto molte volte non è sempre possibile trovare e applicare un vocabolario controllato a tutte le situazioni. Gli svantaggi invece sono: non c'è una

maggior trovabilità degli elementi, molto spesso viene ignorata l'importanza del contesto, e la mancanza di un vocabolario controllato può creare diversi problemi legati alla semantica, all'omonimia (stesso tag per concetti diversi), polisemia, desinenze. Per risolvere alcuni limiti è possibile suggerire all'utente i tag da inserire da una lista di popular tag o da una lista di tag raccomandati, oppure visualizzando una lista di termini le cui lettere iniziali corrispondono ai caratteri inseriti dagli utenti.

4.4.5 Creare sistemi organizzativi coesi

Usare schemi organizzativi esatti se l'utente sa quello che sta cercando, altrimenti utilizzare schemi ambigui, che sono migliori per la navigazione e l'apprendimento associativo. Quando è possibile è opportuno utilizzare entrambi i tipi di schema.

5 Sistemi di navigazione

I sistemi di navigazione devono fornire all'utente il contesto, devono permettere all'utente di spostarsi in modo flessibile tra le unità informative. È necessario bilanciare la flessibilità e la fornitura di troppe opzioni.

5.1 Il design di sistemi di navigazione

L'utente si sposta attraverso le unità informative di un sito web in due modi:

- Browsing (selezione di elementi ipertestuali presenti nella presentazione di partenza, organizzati in gruppi e facenti riferimento alle strutture organizzative del sito)
- Searching (navigazione attraverso un'interrogazione del sistema, svolta attraverso la compilazione di un modulo e la selezione di una delle unità informative proposte come risposta)

Integrare questi due meccanismi permette ad un utente di navigare nel sito attraverso una modalità mista.

Molti browser inoltre forniscono un supporto alla navigazione (aprire un url per accesso diretto, bookmarks, uso di colore nei link ipertestuali, tasti avanti e indietro). Non bisogna danneggiare questo sistema di navigazione (modificare colore link, rimuovere sottolineatura dei link, nascondere l'url di destinazione)

5.2 Costruire il contesto

È necessario costruire quindi un contesto, ad esempio includendo in tutte le pagine il nome dell'ente/azienda oppure un logo e presentando in tutte le pagine la struttura dell'informazione gerarchica e la posizione dell'utente rispetto ad essa.

5.3 Navigation Stress Test

È utile per verificare la capacità di un sito di rappresentare il contesto. Si parte da una pagina a caso, e si rispondono alle domande:

- So dove sono?
- In che sezione principale del sito mi trovo?
- Qual'è la pagina padre di questa pagina?
- Dove mi condurrà la pagina in cui mi trovo?
- I link suggeriscono la destinazione?
- I link sono abbastanza differenziati per aiutarmi a scegliere un link piuttosto che un altro?

5.4 Tipologie di Sistemi di Browsing

Un sito complesso può includere diversi sistemi di navigazione.

5.4.1 Sistema Gerarchico

Gerarchia informativa, permette la navigazione da elementi padre ad elementi figlio, e viceversa.

5.4.2 Sistema Globale

Permette la navigazione verticale (da elementi padre a figlio, anche a passi maggiori di 1) e la navigazione orizzontale (tra elementi dello stesso padre).

5.4.3 Sistemi Locali

Sistema di navigazione utilizzato in un sottosito, cioè in una sezione del sito nella quale le unità informative sono legate da relazioni peculiari (es. catalogo online di un'azienda). È importante estendere il sistema Globale anche all'interno di sottositi.

5.4.4 Sistemi Ad hoc - Trasversali - Contestuali

Le relazioni tra contenuti informativi non riconducibili ad attraversamenti verticali o orizzontali della gerarchia, ad esempio i links contenuti nel corpo di un paragrafo (embedded links) oppure links evidenziati separatamente del layout della pagina.

5.5 Elementi di browsing integrati

5.5.1 Menu di navigazione

È una collezione di links ipertestuali, sottoforma di testo, imagemap o immagini separate.

La progettazione dell'architettura può influenzare la scelta dell'implementazione. Ad esempio, siti con aree in espansione o variabili scelgono soluzioni testuali. Per la navigazione globale è opportuno usare una barra grafica, per la navigazione locale una barra testuale.

I menu permettono di effettuare la navigazione orizzontale all'interno del sito.

5.5.2 Shortcuts

Sono elementi di navigazione che risiedono tipicamente nella home page di un sito: possono essere singoli oppure utilizzati in gruppi.

Permettono all'utente un accesso diretto a elementi che si trovano nei livelli più profondi della gerarchia informativa del sito (sono una navigazione verticale).

Sono associati all'esigenza di rendere noto all'utente l'esistenza di un elemento o di un gruppo di elementi informativi.

5.5.3 Menu pop-up e pull-down

Permettono di rappresentare molti link in forma compatta, nascondono opzioni. Vengono utilizzati per effettuare selezioni secondarie oppure in contesti informativi molto affollati. È sconsigliato utilizzarli per le opzioni informative principali nei sistemi desktop (nei layout per smartphone invece è possibile utilizzarli per ottimizzare lo spazio).

È opportuno che la riga di menù visibile indichi la tipologia dell'informazione.

È possibile anche specificare eventi che possono attivare l'apertura del menu (ad esempio uno swipe, la pressione di un tasto, al passaggio del mouse).

5.5.4 Drawer

I drawer permettono di rappresentare links in forma compatta: può essere utile utilizzarli nei casi in cui sia necessario un livello di strutturazione elevato dei menu pop-up e pull-down.

Vengono utilizzati nei dispositivi con uno schermo piccolo.

5.6 Sistemi di browsing

- Navigazione gerarchica (può essere molto limitante)
- Navigazione globale
- Navigazione locale

- Navigazione trasversale
 - Embedded links (navigazione adhoc con link nel testo)
 - Navigazione trasversale
 - See also

5.7 Elementi di browsing remoti

Sono elementi che forniscono un punto di vista esterno alla gerarchia base del sito. Sono utilizzati come complemento agli elementi di navigazione integrati.

5.7.1 Indice dei contenuti

Rappresenta i livelli della gerarchia informativa, l'indice attivo (che impiega link ipertestuali) facilita l'accesso non lineare ai contenuti.

È utile utilizzarlo quando il sito ha un'organizzazione gerarchica.

5.7.2 Indice analitico

Alternativa per siti che non hanno una forte organizzazione gerarchica.

È utile per gli utenti che sanno quello che stanno cercando; gli elementi dell'indice devono puntare a componenti del sito dove prevalgono i contenuti piuttosto che a componenti dove prevalgono le strutture di navigazione.

5.7.3 Mappa del sito

È una rappresentazione grafica dell'architettura di un sito.

La mappa evidenzia le relazioni tra le componenti informative.

Deve rappresentare gli elementi informativi importanti e le loro relazioni in modo chiaro e significativo.

Le mappe non si prestano ad una lettura TTSS, è preferibile quindi affiancare una mappa ad un indice dei contenuti.

5.7.4 Tour guidato

Utile per introdurre ai nuovi utenti i contenuti del sito. Dovrebbe avere una navigazione lineare ed essere eseguibile dalla home page del sito.

5.8 Sistemi di browsing personalizzati

Possiamo distinguerli in:

- Adattivi: le opportunità di navigazione vengono selezionate dal sistema in base ad alcuni parametri, come il profilo dell'utente o lo storico di utilizzo.
- Adattabili: l'utente può scegliere diverse opzioni di navigazioni differenziate.

5.9 Sistemi di navigazione sociale

Permettono all'utente di navigare il sito utilizzando l'attività di tagging degli altri utenti, si costruisce quindi una folksonomia in modo progressivo. Può essere l'unico sistema di navigazione oppure affiancarne un altro.

5.9.1 Tag clouds

Una folksonomia è rappresentata da una tagcloud, che esprime la frequenza di utilizzo dei tag attraverso la dimensione del carattere tipografico. Può essere costruita a partire da una folksonomia allargata che ristretta.

5.9.2 Interfaccia per l'inserimento di tag

Oltre all'interfaccia di navigazione può essere presente l'interfaccia di inserimento dei tags (nei sistemi di tipo bottom-up l'utente può contribuire ad arricchire la classificazione).

5.9.3 Integrazione nel layout delle interfacce di navigazione sociale

Le interfacce dei sistemi di navigazione possono essere presenti nel layout del sito specifico, nel layout di un altro sito che funge da collettore di tagging degli utenti sul web, oppure secondo modalità miste.

5.9.4 Utenti, risorse e annotazioni

1. Annotazioni top-down: i singoli siti forniscono all'utente fornitore di contenuti le interfacce per inserire risorse e annotarle. L'utente che gestisce la singola risorsa coincide con l'utente autorizzato ad annotare la risorsa (solo lui può annotare la risorsa). Abbiamo una integrazione totale dell'interfaccia di annotazione nel layout.
2. Annotazioni bottom-up tipo 1: le risorse informative e le annotazioni risiedono in unico repository. Le interfacce e le risorse per annotarle sono fornite da un gestore unico.
Abbiamo comunque un'integrazione totale delle interfacce di annotazione nel layout di presentazione, ma in questo caso è consentito ad utenti diversi l'annotazione di risorse gestite da altri.
3. Annotazioni bottom-up tipo 2: le risorse informative e le annotazioni risiedono in siti diversi. È difficile quindi un'integrazione delle interfacce di annotazione nel layout di presentazione delle risorse informative.

5.9.5 Fake tagcloud

Esistono anche della tagcloud fake: in questo caso i contenuti della tagcloud, e la grandezza delle parole è una scelta redazionale, per mettere in risalto alcuni contenuti. In questo caso, non possiamo parlare di tagging sociale.

6 Sistemi di ricerca

6.1 Quando implementare un sistema di ricerca?

- Quando abbiamo molta informazione da percorrere attraverso sistemi di browsing
- Quando il sito è costituito da componenti frammentati
- Quando dobbiamo imparare dagli utenti (analisi search-logs)
- Quando gli utenti si aspettano che ci sia
- Quando il sito è dinamico

6.2 Zone di ricerca

Non è sempre opportuno indicizzare tutto il sito.

Creare search-zones riduce l'effetto apple-and-oranges, ossia la ricerca effettuata su aree eterogenee di contenuti che può essere un ostacolo al ritrovamento dell'informazione.

Le zone di ricerca sono parti di sito che sono state indicizzate separatamente.

Si possono creare suddividendo logicamente o fisicamente i documenti. È possibile crearla a partire da schemi e strutture organizzative.

Con le zone di ricerca però andiamo ad aggiungere un'ulteriore complessità alla ricerca.

6.3 Pagine di destinazione e navigazione

Le pagine di destinazione contengono il contenuto vero e proprio.

Le pagine di navigazione hanno lo scopo di condurre gli utenti verso le pagine di destinazione. Non vengono indicizzate.

Spesso, le pagine appartengono ad entrambe le categorie.

6.4 Componenti di una pagina

La nascita dei sistemi CMS (Content Management System) rende più facile l'identificazione degli elementi della pagina (contenuto informativo, menu di navigazione, pubblicità, disclaimers...) e la loro inclusione o meno dall'indicizzazione. Usare un CMS permette di selezionare gli elementi da indicizzare ad un livello di granularità più accurato rispetto a quello della pagina.

6.5 Presentare i risultati

6.5.1 Quali?

- mostrare come risultato all'utente che sa quello che sta cercando meno informazione rispetto a quella presentata all'utente che non sa quello che sta cercando.

- Agli utenti del primo caso possono bastare alcuni elementi rappresentativi del contenuto.
- La seconda categoria di utenti richiede contenuti più descrittivi.
- In alcuni casi, l'utente può scegliere cosa visualizzare.

6.5.2 Quanti?

Dipende da due fattori, dalla quantità di componenti informativi per componente e dalla caratteristica del dispositivo.

Si parte dalla presentazione di un gruppo limitato di documenti (10 per pagina) e si lascia la libertà di configurare al sistema.

È opportuno visualizzare il numero complessivo di documenti ritrovati e fornire un sistema di browsing per permettere una rapida navigazione tra i risultati.

6.5.3 Elencare i risultati

Esistono due metodi per elencare i risultati.

Sorting: i risultati vengono ordinati secondo un criterio basato su una delle componenti dei documenti (utile quando l'utente deve comparare due o più documenti).

Ranking: risultati ordinati con criteri come la rilevanza del documento, dedotta analizzandone i contenuti (considerando quanti termini della query sono presenti nel documento e quante volte, la vicinanza e la locazione di questi termini) o la sua popolarità. (ad esempio PageRank di Google). Oppure può essere realizzato un ranking attraverso il giudizio di esperti, i quali valutano il valore dell'informazione, o attraverso un modello pay-for-placement.

6.5.4 Raggruppare i risultati

Una tecnica alternativa (o complementare) al Sorting e al Ranking consiste nel raggruppare i risultati in base a qualche aspetto comune.

Come possiamo raggruppare i risultati? Attraverso l'utilizzo di metadati (tipo di documento, data creazione), attraverso l'uso di metadati applicati manualmente (audience, contenuto, tags). Alcuni tool cercano di derivare autonomamente alcuni cluster di contenuti considerando i contesti possibili a partire da una keyword inserita dall'utente.

6.5.5 Esportare i risultati

Un utile complemento può essere quello di esportare i risultati (stampare, inviare via mail, salvare).

Quando si vuole salvare un insieme di elementi interessanti è utile fornire una selezione di un sottoinsieme di risultati.

Il salvataggio di una ricerca, inoltre, è un'alternativa utile per i domini dinamici, per i quali la riesecuzione dell'interrogazione in un altro momento può portare nuovi risultati.

6.6 L'interfaccia di ricerca

È necessario mantenere l'interfaccia di ricerca il più semplice possibile. L'utente non deve utilizzare operatori logici, non deve preoccuparsi di utilizzare un vocabolario controllato e la ricerca viene fatta sull'intero sito.

Per educare l'utente è possibile creare una pagina di consigli, oppure inserire la possibilità di restringere la ricerca ad una search zone o inserendo un collegamento ad un modulo di ricerca avanzata.

6.7 L'interfaccia per i risultati

È necessario fornire un supporto alla revisione, ad esempio descrivendo gli eventuali filtri applicati, mostrando gli operatori booleani applicati implicitamente alla ricerca, mostrando le caratteristiche utilizzate per la presentazione (ad esempio il tipo di sorting) ed evidenziando il numero dei risultati ottenuti dalla ricerca.

6.8 Interfacce di ricerca basate sulla classificazione sociale

In alcune situazioni sono disponibili sistemi di ricerca dei tags.

Succede per le situazioni in cui i tags costituiscono il sistema prevalente di classificazione dell'informazione o nei sistemi caratterizzati dalla presenza di molti tags associati alle risorse.

È importante rendere consapevole l'utenza che la ricerca delle informazioni avviene sull'insieme dei tags e non sull'insieme delle parole che compongono la pagina web.

6.9 Trovabilità

La trovabilità è definita come $Finding = Searching + Browsing$. Per aumentare la trovabilità degli elementi è necessario quindi integrare le due modalità di navigazione di un sito, rendendo possibile passare da una navigazione ad un'altra in modo semplice.

7 HTML

Formato non proprietario basato su SGML e può essere creato e manipolato da una vasta gamma di tool.

SGML è un sistema di regole per definire linguaggi basati su marcatori. Stabilisce il sistema di descrivere documenti in termini di struttura, indipendentemente dall'apparenza.

HTML utilizza un sottoinsieme delle caratteristiche di SGML. Gli elementi di una pagina web sono identificati da marcatori (tag) che danno istruzioni al browser sul ruolo dell'elemento e sulla modalità di visualizzazione del contenuto.

HTML si basa sul principio di mantenere separata l'informazione relativa allo stile dalla struttura del documento.

7.1 Timeline di HTML

- HTML
- HTML 2.0 (1995)
- HTML 3.2 (1997)
- HTML 4.0 (1997)
- HTML 4.01
 - XHTML 1.0
 - XHTML Basic (browser web per palmari)
 - XHTML 1.1 (basato sul framework di modularizzazione)
 - XHTML Mobile Profiler 1.0 (sviluppato da Nokia, per mobile phone)
 - XHTML 2.0
- HTML 5

Mentre XHTML 2.0 (evoluzione di XHTML 1.1) è XML puro, HTML 5 rimane compatibile con gli standard precedenti e fornisce estensioni per lo sviluppo di applicazioni.

7.2 Tag HTML

Un tag X/HTML è racchiuso tra `<` e `>`. Il tag è formato dal nome dell'elemento seguito da una lista (opzionale) di attributi. Abbiamo due tipologie di tag:

- Elementi non vuoti: `< elemento > testo < /elememnto >`
- Elementi vuoti: `< elemento/ >`

Lo standard XHTML richiede sempre il tag finale, opzionale in HTML. Se il contenuto non è presente, va sempre utilizzata la forma con tag iniziale e tag finale per compatibilità con i browser HTML.

7.2.1 Elementi vuoti

Alcuni tag non richiedono il tag finale. Va inserito il carattere `/` prima della fine del tag.

È opportuno inserire uno spazio prima del carattere `/` per compatibilità con i vecchi browser.

7.2.2 Nidificazione di Tag

Un tag può essere inserito all'interno di un altro tag. Non è possibile sovrapporre i tag.

Documenti XHTML sono ben formati se tutti gli elementi hanno il tag di chiusura e sono annidati correttamente.

7.2.3 Attributi di tag

Gli attributi vengono aggiunti per estendere o modificare le azioni di un tag.

Va inserito sempre nel tag iniziale, è possibile inserire molti attributi, separati da spazio. Non è importante l'ordine.

Tutti i tag e attributi devono essere scritti con lettere minuscolo e i valori degli attributi devono essere racchiusi tra doppi apici (richiesto in XHTML).

I valori degli attributi possono venire scritti sia maiuscoli che minuscoli.

In HTML 5, esiste l'attributo booleano, con una sintassi semplificata. La presenza dell'attributo indica un valore true, la sua assenza false.

Se dobbiamo utilizzare caratteri speciali (<, > e " ",) come parte del valore di un attributo possiamo usare stringhe alternative dette CER (es. ", <).

7.3 Caratteri ignorati

In una pagina HTML vengono ignorati:

- Interruzioni di linea
- Tabulatori e spazi multipli
- Tag < p > nidificati
- Tag sconosciuti

7.4 Struttura di un documento XHTML 1

Un documento XHTML 1.0 è composto da:

- Prologo XML (opzionale)
- Dichiarazione del tipo di documento (<!DOCTYPE >)
- Sezione di testa (< head >)
- Corpo del documento (< body >)

L'head e il body devono comparire all'interno del tag < html >.

7.4.1 Prologo XML

Serve a specificare la versione XML e la codifica dei caratteri.

W3C ne raccomanda l'utilizzo, ma alcuni browser non lo gestiscono correttamente. Per inserire la codifica dei caratteri al di fuori del prologo è possibile inserire un elemento content-type nella sezione head del documento.

7.5 Specifica dei caratteri

È necessario stabilire una relazione tra i singoli caratteri utilizzati e il codice numerico che li rappresenta.

Diversi standard effettuano questa mappatura:

- Famiglia ISO-8859: set di caratteri multilingua a 8 bit per scrivere in lingue alfabetiche (ad esempio, Latin-1, Latin-2).
Ha una scarsa portabilità del testo al di fuori di ambiti regionali, ma anche tra sistemi operativi diversi (nei primi anni del Web Windows e MacOS utilizzavano standard proprietari)
- Unicode: standard universale, codifica tutti i caratteri utilizzati nei linguaggi scritti utilizzati in tutto il mondo.

È inoltre possibile, per identificare un carattere, utilizzare una stringa CER (`&#nnnn`, dove `nnnn` sono i riferimenti numeri ad un elemento del set di caratteri espressi in forma decimale o esadecimale (in quest'ultimo caso, prima del riferimento, va inserita una `x`, `xnnnn`)). Oltre ad usare un codice numerico, è possibile utilizzare un nome (`&name`).

Le nuove versioni dei Sistemi Operativi supportano unicode, rendendo quindi inutile le CER. CER comunque permette una corretta compatibilità verso i sistemi meno recenti.

7.6 DTD (Document Type Definition)

Un documento X/HTML valido deve iniziare con una dichiarazione nella quale si dice quale versione di X/HTML viene utilizzata.

Ogni versione è definita da una DTD, che ne descrive in un linguaggio preciso e leggibile da computer la sintassi e la grammatica permesse nel documento.

I DTD sono disponibili online.

Quando il documento XHTML viene esaminato da un validatore, il contenuto viene validato rispetto ad DTD indicato.

HTML 4 E XHTML 1.0 fanno riferimento a 3 diverse DTD che includono 3 diversi sottoinsiemi di tag consentiti nel documento (questo perchè questi formati sono formati di transizione):

- STRICT: codice XHTML privo di tag o attributi di presentazione
- TRANSITIONAL: documento con tag di presentazione
- FRAMESET: da utilizzare se il documento utilizza, oltre ai tag previsti dal TRANSITIONAL, anche le frames.

7.7 Il tag `<html>`

Contiene una dichiarazione di namespace.

Un namespace fornisce un metodo per qualificare elementi ed attributi utilizzati

da un documento XML, associandoli con i namespaces identificati da un riferimento URI.

Per documenti che fanno riferimento a vocabolari multipli, viene descritto un meccanismo che assegna nomi espansi ed attributi per ottenere un'identificazione unifica, ad esempio usando i attributi `xml:lang` e `lang` per specificare la versione xml e la lingua del documento.

7.8 Conformità

Un documento XHTML è conforme se:

- l'elemento radice è `head`.
- l'elemento radice deve indicare lo spazio dei nomi di XHTML usando l'attributo `xmlns`.
- prima della radice deve esserci una dichiarazione di DOCTYPE, il quale conterrà un'identificatore che si deve riferire ad una delle 3 DTD (`strict`, `transitional`, `frameset`).
- deve essere validato rispetto ad una delle 3 dtd.

7.9 Il tag `< head >`

Non ha attributi, contiene tag utilizzati per definire i contenuti del documento:

- `< title >`: descrizione contenuti della pagina. Elemento importante per l'indicizzazione. Obbligatorio in XHTML 1.0. Solo caratteri ascii.
- `< base >`: localizzazione base sul server web, da utilizzare come riferimento per i links presenti nel documento.
- `< link >`: definisce la relazione con un altro documento (usato ad es. per collegare il documento ad un foglio di stile)
- `< script >`

7.9.1 Il tag `< meta >`

Il tag meta è presente in head, ed è utilizzato per molti scopi. I dati inclusi nel tag sono utili per server, browser, e sono invisibili al navigatore.

È necessario specificare l'attributo `content`, per fornire un valore all'informazione. Ci sono due categorie di tag meta:

- `http-equiv`: viene processata come se provenisse da un header di risposta http. Fornisce di conseguenza informazioni che condizionano le modalità con le quali il browser manipola il documento.
L'attributo `http-equiv` con valore `content-type` permette di specificare il set di caratteri utilizzati.

- name: permette di inserire informazioni riguardanti il documento (descriptions, keywords, author, robots (per prevenire l'indicizzazione), rating...).

7.10 Il tag `< body >`

Contiene le informazioni che verranno presentata all'utente sulla pagina. Nel passato sono stati definiti un insieme di tag di presentazione, ora deprecati.

7.11 HTML 5

7.11.1 Struttura di un documento HTML 5

```
<!DOCTYPE html>

<html>

    <head>

        <meta charset="UTF-8">

        <title>Page Title</title>

    </head>

    <body>

    </body>

</html>
```

7.11.2 Doctype

La dichiarazione di tipo del documento viene semplificata: HTML 5 supporta il contenuto esistente di HTML 4.01 o di XHTML 1.0, e le versioni successive dovranno supportare il contenuto di HTML 5. Quindi non è necessario specificare la versione di HTML.

7.11.3 Codifica dei tag e attributi

Puoi scrivere come vuoi, puoi scegliere di mettere le virgolette o meno, di scrivere i tag in maiuscolo...

7.11.4 Obsolescenza

Non ci sono attributi deprecati, ma solo obsoleti, perchè si vuole mantenere la compatibilità con il passato.

Un documento con elementi o attributi obsoleti sarà non conforme.

Tuttavia, alcuni elementi di presentazione sono stati ridefiniti per superare i limiti passati.