



Università
Ca' Foscari
Venezia

Corso di Laurea magistrale in Economia e Gestione delle Arti e delle Attività Culturali

Tesi di Laurea

Applicazione di tecniche SEO ai siti museali Il caso pratico del Museo della Città di Rimini

—
Ca' Foscari
Dorsoduro 3246
30123 Venezia

Relatore

Ch.mo Prof. Renzo Orsini

Correlatore

Ch.mo Prof. Bruno Bernardi

Laureanda

Silvia Lanza

Matricola 836422

Anno Accademico

2012 / 2013

INDICE

INTRODUZIONE	p. 5
I MOTORI DI RICERCA	
1. Funzionamento di un motore di ricerca	p. 11
1.1 Sistema di scansione	p. 11
1.2 Repository	p. 12
1.3 Sistema di indicizzazione	p. 13
1.4 Sistema di ricerca	p. 17
1.5 Sistema di presentazione	p. 18
1.6 Sistema di frontend	p. 19
2. Google	
2.1 Storia di Google	p. 20
2.2 Struttura di Google	p. 21
2.3 I profitti di Google	p. 30
2.4 Google Ranking Factors	p. 32
ANALISI DELLE TECNICHE SEO	
1. L'importanza delle tecniche SEO	p. 39
2. Le principali tecniche di Search Engine Optimization	p. 41
2.1. Analisi di mercato	p. 42
2.2. Analisi interna (on-page optimization)	p. 46
2.3. Analisi esterna (off-page optimization)	p. 60
IL SITO DEL MUSEO DELLA CITTA' DI RIMINI	
1. Analisi di mercato	p. 67
1.1. Le parole chiave	p. 68
1.2. Analisi degli utenti	p. 70
2. Analisi interna	p. 71
2.1. Il nome del dominio	p. 71
2.2. Keyword density	p. 72
2.3. Meta tag	p. 73

2.4.	Title tag	p. 74
2.5.	Meta description	p. 74
2.6.	Rich snippet	p. 75
2.7.	Keywords meta tag e language meta tag	p. 76
2.8.	Robots meta tag	p. 76
2.9.	Heading tag	p. 76
2.10	Ottimizzazione della pagina	p. 77
2.11	Ottimizzazione delle immagini	p. 81
2.12.	Scrittura delle URL	p. 82
2.13.	Google sitelink	p. 83
2.14.	Creazione di una sitemap	p. 84
2.15.	SEO mobile	p. 84
3.	Analisi esterna	p. 85
3.1.	Link popularity	p. 85
3.2.	Link building	p. 87

LA VALUTAZIONE DELL'INVESTIMENTO

1.	Le azioni da effettuare con più urgenza	p. 97
2.	Influenza dei fattori SEO sul traffico d'accesso	p. 99
3.	Il legame tra user online e visitatore museale	p. 100
4.	L'investimento nelle tecniche SEO	p. 102

CONCLUSIONI	p. 104
-------------	--------

INTRODUZIONE

Lo scopo di questa tesi di laurea è lo studio e l'analisi delle principali tecniche di Search Engine Optimization, oltre alla successiva ipotetica applicazione al sito dei Musei Comunali di Rimini¹. Tramite questo elaborato si vogliono mostrare alcune semplici tecniche che possono notevolmente aumentare la visibilità di un sito web e, conseguentemente, aumentarne il traffico d'accesso.

La scelta di applicare queste numerose tecniche alla pagina Web dei Musei Comunali della città di Rimini ha numerose motivazioni. Anzitutto è desiderio di chi scrive quello di aumentare la visibilità e popolarità del sito e delle collezioni ivi contenute che spesso, purtroppo, non vengono sufficientemente pubblicizzate. Abitualmente queste nuove tecnologie dell'informazione e comunicazione vengono applicate nelle imprese. A questo proposito si rimanda allo studio svolto dall'Istituto Nazionale di Statistica nel dicembre 2012², il quale mostra come circa il 96% delle imprese con almeno 10 addetti disponga di una connessione Internet, anche se solo il 64,5% di esse disponga di un sito Web e, ancora meno, il 10,6% offra la possibilità di acquisire servizi/effettuare prenotazioni direttamente tramite la piattaforma online.

I dati, pur mostrando un miglioramento rispetto l'anno precedente (trend positivo che, si suppone, sia proseguito in questa direzione anche durante l'anno in corso) rispecchiano l'andamento più generale europeo, per cui, secondo i dati forniti da Eurostat³, nell'anno 2012, 7 aziende su 10 utilizzano un sito Web, ma solo 2 su 10 lo utilizzano per vendere direttamente prodotti e servizi agli utenti.

A partire dagli ultimi 10 anni anche nel settore culturale, e più precisamente museale, sono state applicate sempre più tecniche ICT, e attualmente quasi tutti i musei dispongono di un sito Web. Spesso le informazioni contenute sono meramente di carattere informativo, più raramente sono di supporto alla segreteria, consentendo di effettuare prenotazioni e acquistare *merchandise* online. Ad ogni modo le energie e il denaro spesi in questo settore, spesso costituiscono una minima parte del budget investito, a favore di forme molto più tradizionali di pubblicità e diffusione.

È ipotizzabile spiegare questa reticenza in diverse modalità. Soprattutto nei musei statali, vincere l'inerzia delle grandi organizzazioni e effettuare grossi cambiamenti come un passaggio e una conseguente priorità al Web risulta spesso molto difficile e ostacolato. A preoccupare sono eventuali violazioni di privacy o sicurezza (si pensi alla maggiore diffusione e possibilità di utilizzo di foto di collezioni di proprietà statale, o anche, al timore di rendere pubblica la localizzazione esatta di reperti preziosi). Oltre a ciò si aggiungono i classici dubbi relativi ai costi

1 <http://www.museicomunalirimini.it/>

2 <Http://www.istat.it/it/archivio/77714>

3 http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-12-046/EN/KS-SF-12-046-EN.PDF

e alle difficoltà di messa in campo e di controllo di competenze specializzate, il cui operato spesso è difficile da valutare in quanto mancano gli *skills* di comprensione e conoscenza delle nuove tecniche caratterizzate da ritmi molto rapidi di evoluzione, proprio nella direzione stessa. A questo proposito l'utilizzo di dati numerici spesso non è sufficiente a valutare l'efficacia e l'efficienza derivanti dall'impiego di un'eccellente pagina Web ben pubblicizzata. Si potrebbe anche pensare che la presenza di fasce d'età caratterizzate da rapporti meno spontanei con la tecnologia spesso possa ostacolare il processo, favorendo invece utilizzi più consueti come utilizzo di power point per convegni o di supporti digitali per rappresentazioni; mentre per quanto riguarda il sito internet, si cerca di riprodurre online esattamente la stessa struttura del museo stesso. Si riportano le informazioni più comuni, si mostra il layout strutturale dell'edificio, qualche fotografia per i pezzi da collezione più significativi, i contatti. Ma il lavoro relativo ad Internet per i più finisce qui, non si pensa che si tratta solo di una minima parte dei potenziali visitatori che cerca esattamente il sito del museo in questione. È vero che creare un sito Web costituisce una sorta di vetrina aperta tutte le ore e tutti i giorni dell'anno, però uno degli aspetti più importanti e caratterizzanti di Internet è l'annullamento dei limiti geografici, il sito di un museo di un piccolo paese di provincia, se creato con sapienza e adeguatamente sponsorizzato può apparire tra i primi risultati di ricerca di un utente dall'altra parte del mondo, al pari di musei più grandi o più famosi.

Le piattaforme informatiche consentono di costruire dei rapporti e diventare oggetto di concorrenza globale, in uno scambio continuo con tutti gli utenti. Forse è questo aspetto che può spaventare, se non conosciuto adeguatamente, prima il processo comunicativo era più compartimentalizzato, studiato e mediato, adesso invece la comunicazione deve essere immediata, essenziale e chiara per tutto.

In questa tesi, nonostante l'appetibilità del tema, non si è potuto prendere adeguatamente in considerazione l'argomento dei Social Networks e come il loro utilizzo, ormai ampiamente diffuso globalmente, in ambito museale possa portare una enorme visibilità e affluenza, soprattutto verso la tanto "desiderata" fascia più giovanile, consentendo di inserire il museo nella quotidianità delle vite dei singoli⁴.

Nel caso specifico, come già detto in apertura, si farà riferimento, con un apposito capitolo di applicazione teorica delle tecniche SEO al sito dei Musei Comunali della Città di Rimini. Data la destinazione specifica, ovvero la città di Rimini, è ragionevole supporre che la maggioranza degli

4 A questo proposito si rimanda alla presenza su Facebook di musei quali il British Museum o del MoMA di New York, che quotidianamente pubblicano immagini delle opere contenute coinvolgendo con tematiche diverse e richieste di testimonianze gli utenti, in una forma anche di fidelizzazione e di presenza continua nella vita degli utenti, anche senza trovarsi fisicamente nella struttura, ma a casa o fuori, durante il tempo libero

utenti che effettuerà una ricerca su un motore di ricerca inserendo semplicemente la parola “Rimini”, sarà composta da turisti. Nonostante gli effetti della crisi abbiano afflitto l'intero territorio italiano e parzialmente influenzato il comportamento d'acquisto dei visitatori, la Riviera romagnola rimane comunque una delle mete preferite a livello nazionale, con numeri molto significativi di quasi 16 milioni di turisti.

Tenendo conto dei dati suggeriti dall'Ufficio Statistica della Provincia di Rimini relativi al turismo in Riviera dell'anno 2012 e raccolti sul blog InfoAlberghi5, è evidente il trend positivo registrato con un + 1,8% nel numero di arrivi⁶, conseguentemente parrebbe essenziale, a chi scrive, che l'utente possa venire subito a conoscenza delle offerte culturali a disposizione.



Nel caso in questione, vista l'importanza della città di Rimini come meta di vacanze non solo per turisti italiani, ma anche tantissimi stranieri, sarebbe ideale poter proporre al momento della

5 <http://romagna.info-alberghi.com/rimini/infografica-turismo-riviera-rimini-2012>

6 A fronte di un -1,6% nel numero di notti trascorse nelle varie tipologie di alloggio

ricerca e documentazione sulla città (nel momento successivo alla prenotazione e antecedente alla partenza), le numerose proposte e i festival culturali che ogni anno vengono promossi sul territorio cittadino. Con ogni probabilità, nel momento di scelta della località di villeggiatura, il turista sarà più invogliato a scegliere la meta in grado di proporre anche alternative alla “vita da spiaggia” nel caso di brutto tempo o comunque di svago artistico e culturale presso il centro storico (peraltro ben collegato al versante marittimo) e per questo motivo anche nella fase antecedente alla prenotazione la possibilità di venire a conoscenza dell'offerta culturale è essenziale e potrebbe essere la determinante nella scelta della destinazione turistica.

Il fattore determinante che ha spinto alla creazione di questa tesi è che attualmente, se sul principale motore di ricerca, Google, viene inserita nella finestra di ricerca la parola chiave “Rimini”, il sito dei Musei della Città risulta completamente assente dalle prime tre pagine di risultati, per posizionarsi a metà della quarta.

Ai primi posti di posizioneranno ovviamente la pagina di Wikipedia⁷ relativa alla città e alla sua storia, in seguito il sito di Rimini Turismo⁸ (tramite la home page di questo sito, cliccando su scopri il territorio > arte e cultura > Musei, Gallerie > Museo della Città è possibile raggiungere una scheda informativa sul Museo e tramite essa accedere al sito web ufficiale) e al terzo risultato il sito del Comune di Rimini⁹.

Come si spiegherà e approfondirà meglio nel capitolo relativo alla spiegazione dell'analisi e dell'applicazione delle tecniche SEO, è dimostrata da numerosi studi la grande importanza del posizionamento di un sito Web tra i risultati di ricerca durante una *query*, ovvero una ricerca, svolta da un utente che potrebbe diventare un potenziale cliente, o meglio in questo caso, un visitatore del museo. Difatti il primo risultato di ricerca otterrà il 33% del traffico web e, più in generale, le pagine web collocate nella prima pagina delle SERP¹⁰ riceveranno il 98% di tutto il traffico di ricerca.

È essenziale dunque per ottenere buona visibilità e aumentare il numero di potenziali visitatori che il sito Web dei Musei Comunali della Città di Rimini migliori il suo posizionamento tra i risultati di ricerca, e ciò è quanto si spera di ottenere mediante questa tesi di laurea.

Nel primo capitolo verrà spiegato il funzionamento dei motori di ricerca in generale e, più nello specifico, come funzionino Google e il suo sistema di *ranking* (ovvero in base a quali criteri i

7 <http://it.wikipedia.org/wiki/Rimini>

8 <http://www.riminiturismo.it/505/0/Visitatori.html>

9 <http://www.comune.rimini.it/>

10 L'espressione inglese **Search Engine Results Page** (abbreviata in **SERP**) rappresenta la pagina di risultati dei motori di ricerca e corrisponde all'elenco ordinato dei risultati restituiti da un motore di ricerca in risposta a una keyword query (una ricerca effettuata utilizzando parole chiave)

risultati vengano presentati all'utente in un determinato ordine).

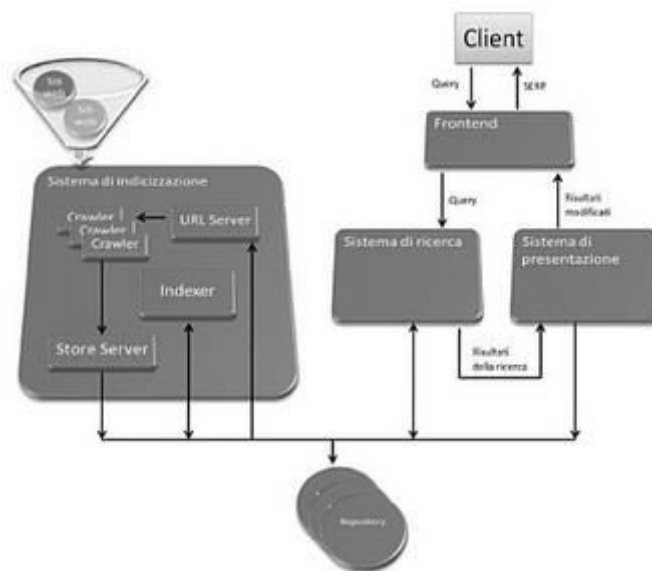
Nel secondo capitolo verranno mostrate e spiegate le principali tecniche di SEO, le quali troveranno poi applicazione teorica nel capitolo successivo.

Infine nel quarto e ultimo capitolo si esploreranno i possibili orizzonti di sviluppo e la direzione di crescita di queste tecniche sempre innovate per garantire ottima visibilità e per rimanere al passo con i nuovi interessi degli utenti e dell'enorme capacità di innovazione dei principali motori di ricerca e strumenti che li utilizzano.

I MOTORI DI RICERCA

Nonostante sul mercato mondiale esistano numerosissimi motori di ricerca, in questa tesi verrà preso a riferimento principalmente Google e questo per via del suo costante incremento di *share* sul mercato Web. Si rimanda al rapporto mensile pubblicato dalla società di metriche online comScore Inc. leader mondiale nella misurazione digitale e analisi del traffico internet, che mostra come nel mese di Maggio 2013 nel *search market* statunitense, dato un volume totale di 20 miliardi di ricerche, il 66.7% (13.353) di esse sia stato condotto usufruendo dei siti di Google, il 17.4% da siti Microsoft (Live e Bing), l'11.9% da Yahoo!, mentre lo *share* rimanente distribuito su altri due network, Ask e AOL¹¹.

Un motore di ricerca è composto da 5 parti fondamentali: il repository, il sistema di indicizzazione, il sistema di ricerca, il sistema di presentazione e il frontend.



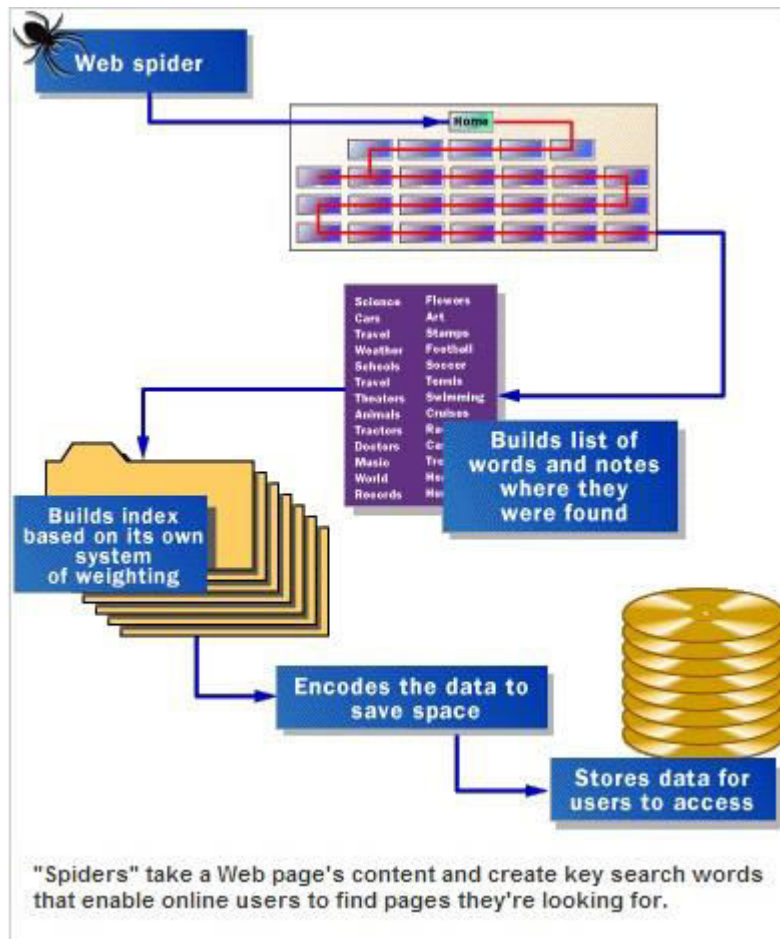
L.Toscano, 2011, pag. 56

¹¹http://www.comscore.com/Insights/Press_Releases/2013/6/comScore_Releases_May_2013_U.S._Search_Engine_Rankings

1. Funzionamento di un motore di ricerca

1.1. Sistema di scansione

Questo elemento del motore di ricerca esegue il *crawling*¹² dei documenti accedendo ai siti Web e ad altri insiemi di documenti (tutti i media file che possono essere indicizzati e recuperati dal *search engine*; per esempio le pagine web o le immagini, i PDF, ecc.). In un secondo momento, verrà eseguita l'identificazione delle singole parole e frasi presenti nei documenti; sarà poi in base



ad esse che i singoli documenti verranno indicizzati.

I web crawlers che eseguiranno la scansione di tutti gli URL inizieranno dalle pagine più popolari e dagli elenchi di server più affollati. Verrà selezionato un sito molto popolare, di cui verranno indicizzate le parole significative presenti su ogni pagina, e poi ogni link trovato sarà aggiunto all'elenco. In questo modo il sistema di scansione si allargherà, proprio come una ragnatela, arrivando, attraverso i link, a coprire le parti più usate del Web.

Una volta che i web crawlers hanno completato il compito di trovare le informazioni sulle pagine Web (questo in realtà è un compito impossibile da terminare, lo stato di perenne evoluzione del Web rende necessaria la continua scansione degli URL da parte di spiders), il motore di ricerca deve memorizzare in maniera utile le informazioni reperite.

¹² I *crawler* sono dei programmi automatici che esplorano il Web metodicamente, onde fornire dati aggiornati ai motori di ricerca. Sono conosciuti con vari nomi, da *crawlers* a *web spiders*, ma anche come *automatic indexers*, anche se la funzione rimane la stessa. Il Web crawling, in sintesi, consiste nell'esplorazione di tutti gli URL dei siti Web che devono essere visitati, i cosiddetti *seeds*, e nell'esplorazione di tutti i *link* contenuti nelle pagine. Gli indirizzi URL di queste pagine vengono visitati più volte in base alle politiche dei singoli motori di ricerca; ciò viene fatto per svolgere un'azione cautelativa contro la trasformazione di alcune pagine aggiunte precedentemente all'indice in *spam*

Di questi documenti memorizzati non verrà valutato il contenuto, ma verrà effettuata una selezione basata su differenti criteri. Alcuni di questi criteri (verranno riportati quelli più diffusi) sono l'esame dei domini presenti all'interno di liste *spam*, il riconoscimento di *spider traps*¹³, ma anche l'esame di eventuali *file robots.txt* o *meta tag*¹⁴. A questi se ne aggiungono anche altri riguardanti la qualità e la consistenza che possono influenzare il reperimento della pagina. In seguito ogni URL verrà controllato periodicamente dal *crawler* onde verificarne eventuali aggiornamenti o modifiche. Nel caso in cui nel tempo trascorso siano avvenute delle modifiche, lo *spider* provvederà in automatico a rivedere le proprie informazioni e ad utilizzarle per tutte le successive fasi di elaborazione.

Si è già spiegato come si muovano i gruppi di *spider*, data l'enorme mole costituita dai miliardi di pagine Web, essi opereranno su "partizioni del Web"¹⁵.

Una partizione può contenere centinaia di siti in relazione al fatto che essi siano ospitati su *host* appartenenti ad una certa classe di indirizzi IP o che soddisfino altre condizioni di appartenenza. Successivamente, mediante uno *Store Server*, i dati letti dagli spider vengono poi compressi e caricati in un *repository*.

1.2. Repository

Il *repository* è gestito da una piattaforma software ad elevate prestazioni, ciò è comprensibile se si pensa alle dimensioni dei dati che un motore di ricerca deve saper conservare e manipolare quotidianamente. Per dare un'idea di questa enorme mole, basti pensare che uno studio¹⁶ pubblicato nel 2008 stimava la quantità di dati processata dai network di computer gestiti da Google come superiore ai 20 PetaByte¹⁷ al giorno.

Un *repository* delle pagine Web ha una struttura tabellare (definita da alcuni studiosi "Webtable"¹⁸) dove ogni riga indica una pagina Web, mentre ogni colonna indica un attributo diverso del documento. Esempi di alcuni attributi possono essere indirizzo URL e HTML della

13 Le *spider traps*, sono delle situazioni particolari in cui i motori di ricerca possono ritrovarsi, quando i loro *crawlers* rimangono intrappolati in un ciclo infinito di codici dove ad una richiesta di *crawling* viene risposto con un'altra richiesta da parte del server, per delle informazioni precise. Questi input per informazioni non possono essere soddisfatti dai *web crawlers*, che si muovono solo per collezionare dati per il motore di ricerca. Per questo motivo le *spider traps* possono mandare in *crash* un server o impedire agli *spider* di tornare su quella pagina

14 Di questi tag si parlerà in seguito, si tratta di strumenti forniti ai *web masters* che consentono loro di evitare l'analisi *crawling* del proprio sito da parte dei motori di ricerca

15 Toscano, 2001, p. 59

16 <http://www.megalab.it/1752/google-gigante-da-20-petabyte-di-dati-al-giorno>,
<http://googleblog.blogspot.it/2008/11/sorting-1pb-with-mapreduce.html>

17 Ovvero più di 20 milioni di Gigabyte in sole 24 ore

18 Toscano, 2001, p. 60

pagina Web, linguaggio, data di scadenza del contenuto, indirizzo IP del server ospitante, ecc. L'indirizzo URL è l'identificatore principale delle pagine Web. Ogni cella della tabella tiene conto di tutti i cambiamenti che possono avvenire nei singoli attributi delle pagine Web. Le versioni memorizzate sono identificate in modo univoco da un "timestamp"¹⁹.

Si possono sintetizzare due elementi chiave: raccoglimento dei dati utili per gli *users*; si tratta delle informazioni contenute nei dati e il metodo in cui esse vengono reperite.

Nel caso più semplice un motore di ricerca potrebbe semplicemente archiviare la singola *keyword* e l'indirizzo URL dove essa è stata trovata. In realtà queste due informazioni non sarebbero sufficienti per un motore di ricerca più sofisticato, non sarebbero in grado di comunicare all'utente se la parola chiave rivestisse un ruolo primario nella pagina, o magari fosse presente solo marginalmente, oppure il numero di volte in cui tale parola venga ripetuta o, ancora più determinante, se nella pagina fossero presenti dei link di collegamento ad altri siti web, contenenti a loro volta, quanto la *keyword* cercata dall'utente. In altre parole, un *repository* di questo tipo non darebbe alcuna possibilità di costruire un elenco, un *ranking*, delle pagine, dalla più importante/u alla meno utile. Per ottenere questo risultato, la maggiore parte dei motori di ricerca immagazzinerà più della semplice parola e l'indirizzo URL dove essa viene reperita. Un *search engine* potrebbe memorizzare il numero di volte in cui questa parola appare nel documento scansionato. Successivamente potrebbe essere assegnato un peso per ogni ripetizione della parola, in base a dove essa appare. Ad esempio se essa è presente nel titolo, nei link o nei *meta tag*, allora avrà valore molto maggiore rispetto ad una singola presenza, in fondo al documento. Ogni motore di ricerca utilizza un sistema di pesi diverso da assegnare ad ogni parola nel suo indice; è per questo motivo che la ricerca effettuata su stesse parole chiave utilizzando diversi *engines* produrrà liste di risultati differenti, con le pagine Web in ordini diversi.

1.3. Sistema di indicizzazione

Tra i vari tipi di indicizzazione si andrà ad approfondire quella più diffusa, consistente nell'identificazione delle frasi nei documenti.

Indicizzazione per frasi

L'identificazione avviene, come già detto, scansionando i contenuti di una pagina Web, ma esattamente come fa un motore di ricerca a riconoscere e selezionare le frasi rilevanti?

Una frase è costituita da una serie di uno o più termini, tende a ripetersi più volte in più documenti, spesso è posta in evidenza rispetto al resto del testo (tramite virgolette, grassetto, ecc.)

¹⁹ Toscano, 2001, p. 60

ed è predittiva²⁰ di altre frasi. Ovvero la sua presenza è indice di presenza di altre frasi. L'identificazione di una frase avviene attraverso più fasi²¹, la prima delle quali riguarda la scansione e l'estrazione delle sequenze di termini. La scansione prevede l'analisi del testo mediante l'utilizzo di una "finestra di scansione" dalla lunghezza predefinita, da minimo 2 a massimo 4-5 termini. Questi termini, come già detto, andranno a costituire le frasi da analizzare. La parola d'apertura della finestra è essa stessa una frase, mentre quelle successive vanno a costituire progressivamente frasi differenti mano a mano che si aggiungono una alla volta alla prima parola.

Nella finestra di scansione non verranno filtrati gli articoli, le congiunzioni o proposizioni e le altre parole di uso comune, così come verranno eliminati i segni di interpunzione.

Una frase viene riconosciuta e segnalata nel caso in cui appaia in almeno 10 documenti, per un numero complessivo di occorrenze maggiore di venti o se il numero di occorrenze rilevanti²² sia maggiore di cinque. Al contrario, verrà ignorata nel caso in cui essa appaia in meno di due documenti, ed il numero di occorrenze rilevanti sia pari a zero²³. In tutti gli altri casi di presenza numerica differente, la frase viene segnalata come possibile.

Le frasi possibili e valide vengono memorizzate (assieme ai loro contatori) in due liste corrispettive separate. Le prime vengono incluse nella lista delle frasi possibili; essa viene comunque utilizzata in fase di applicazione dai criteri di selezione. Anche in questo caso per ogni frase presente vengono memorizzati i contatori onde poi valutare la validità. Nel caso in cui ne venga riconosciuta la validità, la serie di termini viene spostata nella seconda lista, quella delle frasi valide.

Sono solo le frasi appartenenti a questa seconda lista a poter essere elaborate nella fase successiva, quella dell'aggiornamento della matrice di co-occorrenza e, successivamente, nel *pruning*. La matrice di co-occorrenza è di importanza essenziale per l'indicizzazione di una pagina web e si trova nel primo step di questa fase, esso consiste nella determinazione del grado di correlazione tra le frasi.

I numeri delle righe e delle colonne di questa matrice corrispondono al numero di frasi valide identificate. Nel momento in cui viene scansionata e inserita una frase valida nuova, essa verrà aggiunta alla matrice di correlazione. Qualora invece una frase già presente in matrice dovesse

20 Ciò significa che la sua presenza in documenti diversi è indice di una probabile presenza di altre frasi collegate

21 Toscano, 2001, pp. 61-70

22 Le occorrenze rilevanti sono quelle che vengono evidenziate o mediante marcatori di testo (grassetto, ecc.) e grammaticali (virgolette, ecc.) oppure quando la frase viene usata come *anchor text* nei link

23 Questi valori numerici risultano scalati in funzione del numero di documenti che compongono la collezione, in questo caso si assume che l'insieme di pagine analizzate sia di un milione di elementi; nel caso in cui il numero fosse il doppio (compreso tra uno e due milioni) i numeri esposti andrebbero circa raddoppiati

subire modifiche sul web, allora il contenuto della cella verrà aggiornato, senza aggiungere nuove righe. In ogni cella della matrice viene memorizzato un indicatore di co-occorrenza della frase posizionata sulla riga, rispetto a quella posizionata in colonna.

“L'indicatore di co-occorrenza è pari al numero di volte che la frase ff (frase posizionata sulla riga) compare nella finestra secondaria assieme alla frase fk (frase posizionata sulla colonna)”²⁴.

Esso viene identificato da una finestra di scansione detta secondaria, onde distinguerla dalla primaria, utilizzata per riconoscere le frasi. Le due finestre si distinguono sia per la lunghezza (la secondaria può rilevare fino a trenta termini), sia per la direzione di analisi che si muove sia verso i termini di sinistra che verso i termini di destra della parola cercata.

Tramite il *pruning* invece il motore di ricerca riuscirà a tenere solo le frasi predittive di altre frasi valide. Nel secondo step della seconda fase, il motore di ricerca calcolerà il guadagno informativo che esiste tra la frase selezionata ff e tutte le altre frasi valide fk . Così facendo si otterrà un rapporto tra la frequenza di co-occorrenza attuale e quella stimata, se esso è maggiore di zero allora le due frasi sono co-occorrenti. Questo rapporto verrà memorizzato per ogni coppia nella corrispondente cella della matrice di co-occorrenza.

Le frasi non predittive verranno eliminate qualora il guadagno informativo non sia superiore ad una determinata soglia.

Nella fase successive verranno individuate le frasi incomplete, ovvero frasi “in grado di predire solo estensioni di esse stesse, ma non di altre frasi”²⁵. Ad esempio la digitazione della frase incompleta “Mostra di” potrà predire diverse frasi, dette estensioni, come “Mostra di Tiziano a Roma” o “Mostra di Manet a Venezia”. Queste frasi possono essere individuate tramite un cosiddetto *string matching* tra la frase ff digitata e le frasi fk , ottenute calcolando il guadagno informativo (che, si ricorda, deve essere maggiore di una certa soglia). Nel caso in cui esse siano tutte estensioni di ff , allora la frase verrà rimossa dalla lista delle valide e trasferita in quelle delle incomplete. Se esiste però almeno una fk non estensione di ff , allora quest'ultima verrà ritenuta valida.

Data la loro generalità le frasi incomplete devono essere rimosse, ma non eliminate definitivamente; vengono mantenute in una collezione di dati, la “lista delle frasi incomplete”. La loro utilità sorge nel momento in cui l'utente è sul punto di digitare una *query* su Google e il motore, riconoscendo la frase incompleta, può così suggerire le possibili estensioni o procedere direttamente con la ricerca.

24 Toscano, 2009, p. 65

25 Toscano, 2009, p. 67



A questo punto dell'elaborazione, il motore di ricerca sarà in possesso di un ampio elenco di frasi valide, predittive di almeno un'altra frase non costituente un'estensione. Le ultime due fasi riguardano rispettivamente, la selezione delle frasi altamente correlate tra loro e la loro organizzazione in *cluster*.

Per generare la lista ordinata di frasi correlate alla frase ff ricercata dall'utente, sarà necessario calcolare il guadagno informativo tra esse e ottenere un rapporto superiore ad un valore predefinito, generalmente molto alto²⁶. All'interno della lista le frasi verranno ordinate in funzione del rapporto di guadagno effettivo rispetto alla frase ff . Evidentemente, per prime appariranno quelle con i valori maggiori, ovvero quelle più predette dalla frase ricercata. Nell'ultima fase dell'indicizzazione per frasi, quelle più associate tra loro verranno organizzate in *cluster*.

I *cluster* vengono utilizzati durante la ricerca per organizzare i risultati, permettendo di discernere i documenti da includere e in quale ordine presentarli; sostanzialmente consentono di individuare una sorta di rete di relazioni tra frasi valide. Le relazioni possono essere mutuali, per cui una frase predice un'altra frase e viceversa, oppure essere monodirezionali, quando il viceversa non sussiste. È anche possibile che una frase abbia più cluster, sovrapposti tra di loro, e in questo caso vengono distinti in “cluster primario”, “cluster secondario”, ecc. I cluster sono identificati da due tipologie di dati, entrambe memorizzate nella matrice di co-occorrenza²⁷:

- il *cluster bit vector*, che costituisce una rappresentazione binaria del cluster primario, un

²⁶ Se si assume che questo valore possa essere uguale a cento, allora la frequenza di co-occorrenza reale dovrà essere cento volte superiore a quella stimata

²⁷ Toscano, 2009, pp. 69-70

bit 1 per ogni frase appartenente al cluster;

- il *cluster ID*, che è un codice univoco.

Si è già ripetuto più volte che al termine di questo processo (che pare molto lungo e complicato da spiegare, ma che nella realtà impiega poche frazioni di secondo e viene svolto automaticamente) i risultati vengono mostrati indicizzati.

L'indice di presentazione è costituito dall'elenco delle frasi valide trovate e per ognuna di esse viene memorizzata una lista dei documenti (detta *postings list*) da cui sono state tratte. Abbinato ad ogni documento vi è un insieme in informazioni ausiliarie riguardanti l'occorrenza delle frasi correlate a *ff* in ogni documento. Alcune informazioni ausiliarie sono:

- il *contatore* di frasi correlate: è costituito da una sequenza di numeri tali per cui ogni elemento corrisponde al numero di occorrenze nel documento di ogni frase correlata a *ff*;
- il *flag* di frasi correlate: è costituito da una sequenza di numeri tali per cui ogni elemento è una coppia di bit. Il primo bit suggerisce l'occorrenza della frase correlata in un determinato documento, mentre il secondo la presenza di almeno una frase correlata secondaria.

Questa sequenza di bit può essere interpretata come un numero in base 10 che in questo caso prende il nome di indice di rilevanza.

Questo indice sarà tanto più grande quante più frasi correlate a *ff* sono presenti nel documento preso in analisi. In altre parole, l'indice esprime proprio la rilevanza di una frase in un dato documento ed è dato dal numero di frasi in correlazione con quella cercata dall'utente.

1.4. Sistema di ricerca

Si tratta della parte del *search engine* che recupera e fornisce i documenti cercati dall'utente durante una *query*. L'identificazione dei documenti viene effettuata tramite scansione delle frasi inserite nella finestra di ricerca, per ognuna di esse è associata una *posting list* e dall'intersezione di tutte le *lists* si ottengono i documenti ricercati. Questi risultati vengono presentati in ordine di rilevanza. Questo ordine, *ranking*, dei documenti può basarsi o sull'indice di rilevanza o su *scoring*. Nel primo caso verranno visualizzati per primi i documenti con il maggior numero di frasi correlate alla *query*; ciò è consigliabile perché, almeno a livello semantico, i documenti sono sicuramente correlati all'argomento ricercato. In questo caso i documenti, pur contenendo un numero limitato delle parole specifiche inserite dall'utente, contengono elevate correlate e dunque saranno senz'altro rilevanti alla ricerca. Attualmente, i principali motori di ricerca interpretano l'importanza come la popolarità: più un sito è conosciuto, documento o pagina che

sia, maggiori informazioni di valore è probabile che vi saranno contenute. Questo *modus operandi* finora si è sempre dimostrato operare bene a livello pratico, e i motori di ricerca hanno continuato a fornire risultati utilizzando delle metriche in grado di interpretare la popolarità. Rilevanza e popolarità non sono determinate manualmente, al contrario esistono degli algoritmi molto sofisticati che dividono i risultati importanti da quelli non rilevanti e organizzano i primi in ordine di importanza. Nel caso di un ordinamento basato su *scoring*, il punteggio finale viene calcolato come una media degli *score* parziali pesata. Riportando la formula si tratta di calcolare:

$$\text{“Score finale} = 0.30 \cdot (\textit{body hit score}^{28}) + 0,70 \cdot (\textit{anchor hit score}^{29})\text{”}^{30}$$

1.5. Sistema di presentazione

Nella parte finale della ricerca, verrà presentata all'utente una serie di risultati; essi possono essere raggruppati per argomento o alcuni possono venire rimossi, in quanto simili ad altri documenti. Nonostante un buon motore di ricerca possa essere in grado di produrre, per ogni ricerca, centinaia di documenti corrispondenti, pare chiaro che molto difficilmente l'utente esaminerà più di 30/40 documenti alla volta. Dato che, come già spiegato, un search engine, automaticamente organizza i vari documenti in cluster correlati a livello tematico, può esserci il rischio che documenti importanti non vengano visualizzati dall'utente. Infatti se i primi 100 risultati mostrati sono stati ottenuti dai primi raggruppamenti di cluster, e i successivi 100 da altri raggruppamenti, l'utente potrebbe perdersi una buona parte di documenti potenzialmente rilevanti. La soluzione trovata consiste nel poter produrre dei campioni di documenti estratti da ogni cluster in modo da mostrare una equa diversificazione dei contenuti. Ciò può essere realizzato estraendo o un numero fisso di documenti da ogni cluster, per esempio i 10 più rilevanti, oppure un numero proporzionale alla mole dei dati di ricerca, impostabile direttamente dall'utente.

Nell'elenco dei risultati, i motori di ricerca mostrano all'utente il titolo della pagina Web, seguito da un sottotitolo e da una breve descrizione dei contenuti. Questa breve sintesi viene chiamata *snippet*, e il processo creativo che ne determina la creazione è la *summarization*.

Sostanzialmente durante questo processo vengono riportate le frasi correlate alla *query* contenute nel documento e evidenziate, il motore di ricerca li memorizzerà assieme al documento corrispondente.

28 Punteggio determinato dal numero di frasi correlate alla *query* contenute nel documento

29 Punteggio determinato dalla somma degli *inbound score component*, ovvero della percentuale di attinenza di ogni documento trovato alla frase di *query*. Questa percentuale è maggiore quando esistono pagine Web contenenti la frase di *query* e altre frasi ad essa correlate

30 Toscano, 2001, p. 75

Altra caratteristica interessante del sistema di presentazione di un *search engine* è l'individuazione, e possibile rimozione, dei risultati duplicati. Mediante un meccanismo di comparazione dei valori di *hash*³¹ contenuti negli *snippet* dei documenti e di conseguente corrispondenza, il motore di ricerca determina lo status di duplicato di un risultato, al fine di rimuoverlo dalla SERP. Il documento mantenuto è generalmente quello classificatosi con un *ranking* più alto.

1.6. Sistema di frontend

Il sistema di *frontend* è la parte del motore di ricerca che riceve la richiesta da parte dell'utente e la inoltra al sistema di ricerca. In seguito all'elaborazione il *frontend* risponde all'utente presentando in maniera ordinata e formattata i risultati.

2. GOOGLE

Il motore di ricerca di cui si vuole approfondire il funzionamento è Google. Questa scelta, le cui motivazioni sono state già illustrate nell'introduzione, si accompagna alla popolarità del sito, classificato come il più cliccato negli USA e il secondo a livello mondiale, dopo Facebook³². Per poter gestire la mole di dati che Google documenta e immagazzina ogni giorno, sono stati edificati 13 data center³³: 7 nelle Americhe, 3 in Asia e 3 in Europa.

Attualmente il numero esatto di server impiegati da Google non è conosciuto, anche se è possibile rintracciare varie stime. Numerosi studi³⁴ riportano che il motore di ricerca utilizzi circa un milione di server, poco più del 2% dei server mondiali. Google non ha mai rivelato il numero esatto di server in esecuzione nei suoi data center, ma nuove stime più recenti³⁵ (e basate su informazioni dell'azienda condivise con il professore di Stanford Jonathan Koomey³⁶, riguardanti i consumi energetici dei data center) suggeriscono che il numero di server utilizzati si aggiri attorno ai novecentomila. Google ha recentemente rivelato alcune informazioni riguardo alla struttura di un data center (relative ad un singolo edificio, costruito nel 2005) per cui esso conterrebbe più di 45.000 server³⁷. Nel maggio del 2012 Google ha anche pubblicato delle fotografie di alcuni dei

31 Un insieme di semplici periodi viene denominato *sentence*, più *sentence* possono essere unite in una stringa. Questa stringa può avere lunghezza variabile, mediante la funzione non invertibile *hash*, come riportato da Wikipedia, è possibile ottenere un testo di lunghezza fissa predefinita

32 <http://www.alexa.com/siteinfo/google.com> , <http://scripts ranking.com/data/details.aspx?theurl=google.com>

33 <http://www.google.com/about/datacenters/inside/locations/index.html>

34 <http://www.chromeos.eu/2011/08/03/google-ha-900-000-server/>

35 R. Miller, 2011

36 <http://www.koomey.com/post/8323374335>

37 R. Miller, 2009

server, come parte integrante di una grande presentazione riguardante il percorso svolto da una mail attraverso internet, per arrivare al destinatario.

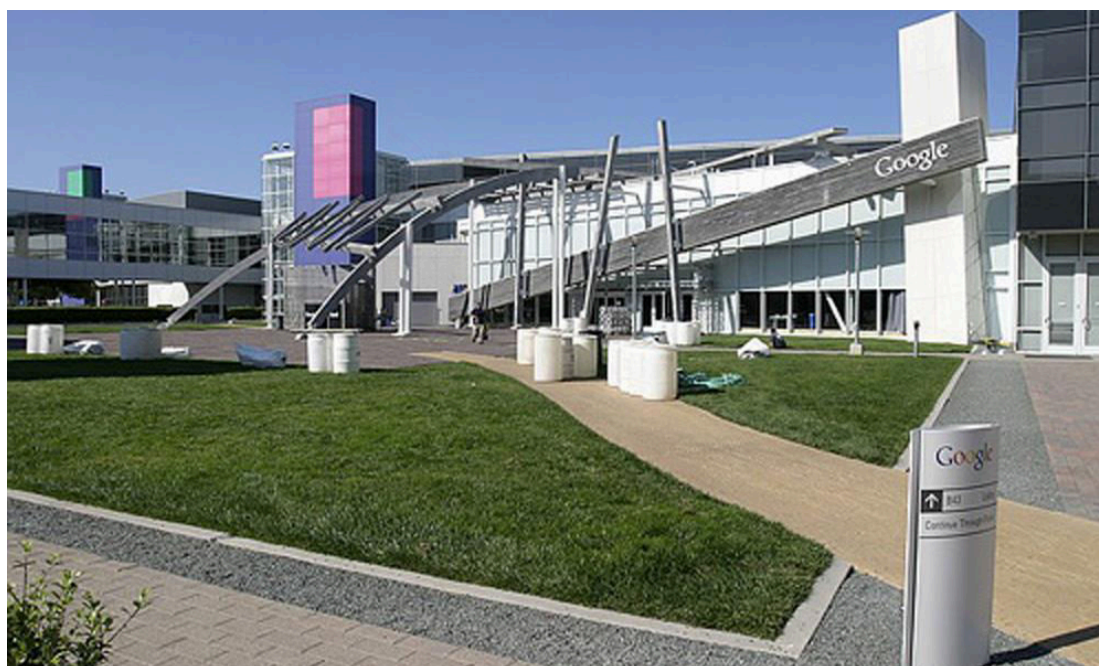


Questa immagine (fonte: Google) mostra file di scaffali carichi di server e immersi nella luce verde dei LED posizionati su ogni cassetto.

2.1. Storia di Google

Per quanto riguarda la storia di questo *search engine*, Google vede i suoi inizi come progetto di ricerca accademico a opera dei due iniziatori Sergey Brin e Lawrence Page, della Stanford University. Il primo progetto risale al 1996, e venne chiamato dai due studenti BackRub. Il motore di ricerca fu creato in modo da essere in grado di analizzare i link di connessione tra le pagine Web per determinare la veridicità e autorità dei siti. Due anni dopo, Page e Brin crearono il loro primo data center, nel dormitorio del primo dei due. Spinti dal collega universitario David Filo³⁸, decisero di avviare la loro impresa e cominciarono a cercare degli investitori. In seguito all'ottenimento di più di un milione di dollari, riuscirono ad avviare il 7 settembre 1998 in California, nel garage di un amico, l'azienda Google, Inc. Riuscirono subito a guadagnarsi una buona reputazione, passando dalle 10,000 ricerche al giorno del primo anno, alle 500,000 di quello successivo e spostandosi dal garage della cittadina di Menlo Park alla nuova struttura nominata Googleplex, a Mountain View.

38 Noto ai più per aver avviato alcuni anni prima il portale web Yahoo!, assieme al co-fondatore Jerry Yang



Il numero di utenti continua ad aumentare, la buona reputazione si accompagna all'affidabilità, effettività e velocità dei risultati. Nel 2000 Google rimpiazza il motore di ricerca interna di Yahoo!, fornendo risultati di ricerca supplementari. Attualmente, con più del 50% di quota del mercato totale di ricerca, Google costituisce una piattaforma di comunicazione e pubblicizzazione essenziale, la prima da conoscere e su cui investire, al fine di far espandere la propria attività.

2.2. Struttura di Google

Per quanto riguarda il funzionamento del motore di ricerca Google, l'azienda stessa³⁹ presenta il proprio operato come diviso in tre frasi principali:

- 1) Scansione e Indicizzazione (Crawling and Indexing)
- 2) Utilizzo di algoritmi (Algorithms)
- 3) Eliminazione dello spam (Fighting Spam)

2.2.1. Il processo di scansione

Googlebot

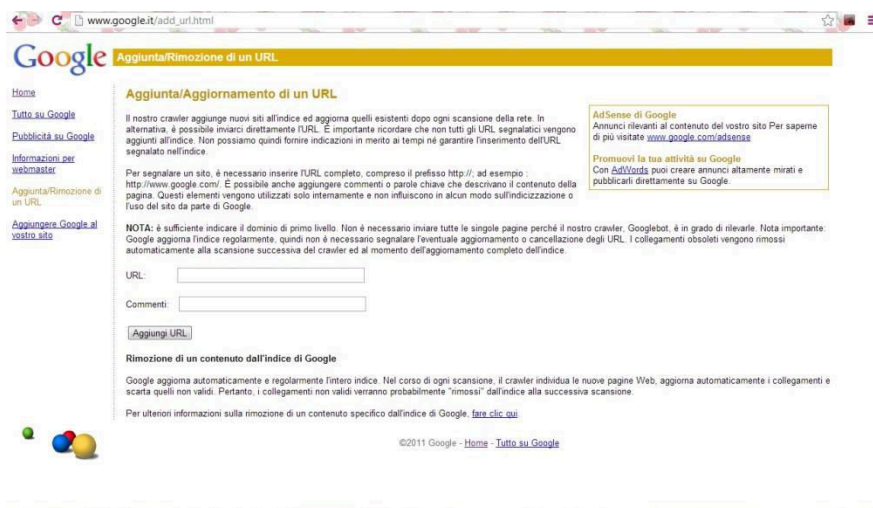
Il protagonista indiscusso della prima fase, in grado di trovare e recuperare le pagine Web, è Googlebot. Si tratta di un *web crawling robot* che, in seguito alla scansione, fornisce le pagine per l'indicizzazione. Googlebot funziona circa come un web browser: invia una richiesta ad un server per trovare una pagina web, una volta trovata la scarica interamente, e infine la consegna all'indice

39 <http://www.google.com/intl/it/insidesearch/howsearchworks/thestory/>

di Google. Questa spiegazione è molto semplificata, difatti Googlebot è costituito da tanti computer che richiedono e recuperano contemporaneamente migliaia di pagine web, molto più velocemente di quanto si possa fare con un semplice web browser .

Googlebot è in grado di recuperare le pagine web in due modi:

- attraverso un modulo Add URL, www.google.com/addurl.html



- trovando collegamenti tramite *crawling* attraverso il web

Il primo metodo ha una complicazione, dovuta al fatto che gli *spammer* sono in grado di creare dei *bot* automatici in grado di “bombardare” continuamente con milioni di URL di propaganda commerciale il modulo Add URL. Per respingere questi tentativi di *spamming*, Google utilizza dei semplici test, come delle lettere distorte/ondulate da inserire in una finestra, prima di poter inserire il proprio URL. Queste serie di lettere o parole non possono essere inseriti automaticamente da sistemi computerizzati, e costituiscono quindi una buona soluzione al problema dell'inserimento di *spam link* (si tratta dei cosiddetti codici CAPTCHA⁴⁰).

Il secondo meccanismo è potenzialmente preferibile in quanto il *crawler* tende a incontrare poco *spam* e ciò si verifica perché quasi tutti i proprietari di siti Web si collegano solo a pagine che ritengono di buona qualità. Quando Googlebot recupera una pagina, automaticamente riordina i link presenti sulla stessa per poi effettuare la scansione. In questo modo, impiegando poco tempo, Googlebot può creare un elenco contenente ampie porzioni del web.

Questa tecnica è conosciuta con il nome di “deep crawling”, sia per via dell'alto numero di pagine che si possono scoprire, sia per la possibilità di analizzarne tutti i link in profondità. Esistono

40 Come citato da Wikipedia, l'acronimo significa “*Completely Automated Public Turing test to tell Computers and Humans Apart*” ovvero, in italiano: “Test di turing pubblico e completamente automatico per distinguere computer e umani”

numerose problematiche che il web crawler di Google deve affrontare, a partire dall'enorme vastità del web e dei tempi di scansione e analisi che possono essere più o meno lunghi. A ciò si aggiungono le imponenti liste di URL che si generano e accumulano di conseguenza e che devono essere analizzate facendo attenzione a evitare doppioni di link nell'indice di Google. Contemporaneamente per garantire risultati sempre aggiornati è necessario che Googlebot controlli periodicamente le pagine, ciò per assicurarsi di annotare in indice eventuali cambiamenti avvenuti in seguito alla prima scansione. Questa necessità viene adempita dal meccanismo noto come "fresh crawling". Difatti, per mantenere l'indice consistente e coerente, Google scansiona continuamente quelle pagine che più sono note per i loro cambiamenti continui; ad esempio le pagine dei quotidiani sono scaricate giornalmente e quelle contenenti gli indici e le quotazioni della borsa ancora più frequentemente. Googlebot fornisce all'indice tutti i contenuti delle pagine che trova, al fine di memorizzarli nel database. La ricerca effettuata per parole chiave svolta da Googlebot è simile a quella effettuata dagli altri motori di ricerca, con i *crawlers* che si muovono da link a link, costruendo mano a mano un indice di *keywords*. Google fa riferimento a questo indice quando un utente effettua una ricerca e mostra l'elenco di pagine risultato contenenti esattamente quanto cercato. Gli *spiders* di Google possono avere anche delle funzioni avanzate, come l'essere in grado di distinguere tra le pagine Web dai contenuti effettivi e i siti reindirizzati, ovvero quelle pagine che esistono solo per dirigere il traffico ad una pagina Web differente. Ad ogni modo, la loro caratteristica più importante è quella di essere in grado non solo di rilevare la parola chiave stessa, ma anche la sua posizione all'interno del testo. Ciò tornerà utile nell'indice, il quale oltre ad essere ordinato alfabeticamente per termine di ricerca, riporta per ogni voce un elenco di documenti, tra cui la keyword e la posizione all'interno del testo in cui essa si verifica. Un sistema di dati così strutturato permette un accesso rapido ai documenti che contengono le parole di ricerca degli utenti. Per migliorare le performance di ricerca, Google ignora (o meglio, non indicizza) le parole più comuni, denominate "stop words"⁴¹. Questo perché esse sono talmente diffuse che servirebbero poco ai fini di una buona ricerca e finirebbero solo per aumentare la massa dell'indice e rallentare i tempi di ricerca; per questi motivi possono essere ignorate. Allo stesso modo vengono ignorati spazi multipli, punteggiatura e le conversioni delle lettere da maiuscole a minuscole. Gli *spiders* di Google operano ad una velocità enorme, alla richiesta fatta ai due creatori Brin e Page di illustrare esattamente quanto velocemente, venne risposto quanto segue. I due spiegarono di aver costruito il loro primo sistema in modo da poter usare più *spiders* contemporaneamente, almeno tre alla volta. Ogni spider poteva mantenere

41 Per *stop words* si intendono gli articoli o le proposizioni

almeno 300 connessioni alla volta con differenti pagine Web. Nei momenti di massima performance, utilizzando 4 spiders, il sistema era in grado di scansionare oltre 100 pagine al secondo, generando circa 600 kilobytes di dati, al secondo. Per mantenere il sistema così rapido, bisognava costruire una struttura in grado di fornire le informazioni necessarie continuamente ai *web crawlers*. Per questo motivo il primo sistema di Google aveva un server dedicato a fornire gli URL agli spider. Sempre per evitare rallentamenti inoltre, il motore di ricerca non si appoggiava a un provider di servizi internet per il DNS (Domain Name Server, server che traduce il nome del server in un indirizzo) ma ne aveva uno proprio.

2.2.2. Il processo di indicizzazione

Scansione dell'indice e utilizzo di Google File System

L'elaborazione della richiesta consiste in due processi separati. Nel primo gli index server scansionano l'indice contenente ciascuna frase della *query*. Il processo è simile a quello già visto nel paragrafo dedicato al funzionamento generico di un *search engine*: le *posting lists* di ogni singola frase inserita vengono incrociate e successivamente viene calcolato il punteggio di correlazione. Questo punteggio di rilevanza sarà una parte dei fattori che determinano l'ordine nella SERP. Il processo di ricerca è molto imponente, le pagine da scansionare, gli indirizzi URL memorizzati nella WebTable⁴² di Google sono numerosissimi, si parla di contenuti dell'ordine di PetaByte. L'unico modo per gestire una mole di questo tipo è suddividere gli *index server* in *index shards* (definiti anche “blocchi contigui”⁴³) contenenti ognuno un sottoinsieme di documenti. Il risultato di questo primo processo è l'ottenimento di un elenco contenente i codici identificatori di tutti i documenti, detti “docID”. Questi codici servono poi nella seconda fase, quando a partire da essi, vengono recuperati a opera dei *document server*, indirizzo URL e *snippet* di ogni documento, per poi infine presentarli all'utente. Al termine di questi due processi il Google File System (GFS) fungendo da frontend, genera la pagina HTML della SERP per l'utente finale.

Google utilizza GFS per organizzare e manipolare enormi files e per permettere agli sviluppatori delle applicazione le risorse di ricerca e sviluppo di cui hanno bisogno. Questo file system è stato ideato nei laboratori di Google e alcuni suoi dettagli rimangono un mistero. Google non intende rivelare quanti computer operano nel GFS, a parte generiche affermazioni di “migliaia di computer”. Nello stesso *paper* pubblicato dall'azienda nel 2003 si legge che GFS è un “file system

42 La webtable di Google è un repository basato su BigTable. Si tratta di un sistema di storage per la gestione dei dati strutturati caratterizzato da alte prestazioni nell'accesso come milioni di read/write al secondo, scalabilità, .. è un layer di servizi costruito su MapReduce e altre soluzioni proprietà di Google

43 Toscano, 2001, p. 81

scalabile distribuito per applicazioni data-intensive”⁴⁴. Esso fornisce tolleranza all'errore durante l'esecuzione su hardware economici ed ha ottime prestazioni ad elevata aggregazione per grandi numero di utenti. Risponde alle necessità di *storage* ed è ampiamente distribuito all'interno di Google come piattaforma di archiviazione per la generazione e l'elaborazione dei dati utilizzati dagli utenti.

2.2.3. Gli algoritmi di Google

Attualmente sembra scontato poter effettuare una ricerca con Google, con qualsiasi mezzo dotato di una connessione al Web, e ottenere migliaia di risultati in pochi secondi. La realtà è che dietro a questo “semplice” e veloce procedimento esistono numerosi meccanismi estremamente sofisticati, ma soprattutto costantemente aggiornati. Come citato direttamente da Google, nelle pagine dedicate alla spiegazione del suo Search Engine: “Per una normale ricerca esistono migliaia, se non milioni, di pagine web con informazioni utili. Gli algoritmi sono i processi informatici e le formule che trasformano le tue domande in risposte. Attualmente gli algoritmi di Google si basano su oltre 200 segnali univoci o "indizi" che consentono di intuire che cosa stai realmente cercando. Questi segnali includono elementi quali i termini presenti nei siti web, l'attualità dei contenuti, l'area geografica e il PageRank”⁴⁵

Come si è già visto, sono i vari algoritmi, seguiti dai numerosi aggiornamenti, di Google che effettuano la ricerca per l'utente, scansionando le pagine Web contenenti le parole chiave inserite nella finestra di ricerca dall'utente e presentandole poi in un determinato ordine. In questi paragrafi si procederà ad illustrare il funzionamento e lo scopo di Google Panda, Google Penguin e infine si analizzerà la metodologia di creazione dell'elenco della pagina dei risultati di ricerca, dai meccanismi di PageRank ai più importanti fattori di ranking che bisogna conoscere.

Google Panda

L'algoritmo attualmente utilizzato da Google è Panda. La prima versione è entrata in vigore nel febbraio 2011 ed è in grado di analizzare la qualità dei siti Web. Nel momento in cui vengono introdotte delle modifiche all'algoritmo di Google, gli esperti SEO appaiono sempre preoccupati da un potenziale sconvolgimento del posizionamento dei siti Web da essi curati. In seguito all'introduzione della prima modifica di Panda, i siti di bassa qualità furono immediatamente penalizzati, mentre quei webmasters che avevano seguito le norme di qualità di Google guadagnarono posizioni rapidamente nella SERP. L'utilizzo di Panda era mirato all'eliminazione dei siti che non avevano sufficiente contenuto di qualità ed erano più orientati all'ottenimento di un

44 <http://research.google.com/archive/gfs.html>

45 <http://www.google.com/intl/it/insidesearch/howsearchworks/algorithms.html>

profitto piuttosto che alla generazione di contenuti utili agli utenti⁴⁶.

Si elencano alcuni fattori essenziali per potersi posizionare ottimamente nella SERP di Google in seguito alle introduzioni delle varie modifiche di Panda:

- Un tempo di caricamento ottimale del sito internet
- Il tempo medio di permanenza da parte degli utenti sulle pagine del sito Web
- La *bounce rate*, ovvero la percentuale degli utenti che abbandonano la pagina Web subito dopo averla aperta
- La percentuale di visitatori che ritornano sulla pagina Web

Tutti questi valori possono essere verificati nella pagina di Google Analytics⁴⁷ gratuitamente.

Quello che pare essere evidente, viste le continue migliorie apportate agli algoritmi di Google, è che l'obiettivo aziendale rimane quello di ottenere un *ranking* sempre più adatto alle ricerche del singolo utente, quasi ad interpretarne le parole chiave e a comprenderne le esatte intenzioni. Ciò potrebbe essere svolto da un interlocutore umano, ma che questo obiettivo possa essere portato avanti da un sistema computerizzato, che dovrebbe interpretare e comprendere la richiesta dell'utente, è la vera sfida che l'azienda sta cercando di vincere. L'implementazione di Panda voleva garantire agli utenti ottime performance sotto 4 aspetti principali:

- La facile comprensione del sito, dalla struttura user-friendly
- Il reperimento di contenuti di pagina originali e esatti rispetto alla *query*
- Evitare eccessivo fenomeno di *advertising* rispetto ai contenuti delle pagine
- La presenza di collegamenti *outbound* del sito, sostanziali e validi

L'ultimo aggiornamento di Panda risale al 18 luglio 2013, ed è stato accompagnato da numerose polemiche, riguardanti il fatto che i siti che pare abbiano avuto la meglio, siano quelli che hanno scelto di utilizzare Google+. Allo stesso modo alcuni siti che avevano subito arretramenti nel loro posizionamento, sono riusciti a guadagnare terreno nel momento in cui hanno scelto di creare un profilo su Google+. Quest'ultimo aggiornamento ad ogni modo sembrerebbe aver avuto un impatto molto più leggero di quelli precedenti, non ci sono state fluttuazioni significative nei risultati⁴⁸.

Google Penguin

Si tratta dell'aggiornamento di Google Panda, lanciato il 26 aprile 2012 sul blog Webmaster di Google, inizialmente solo in Europa (Germania e Gran Bretagna) e poi successivamente negli Stati Uniti. L'algoritmo è stato perfezionato per individuare e punire tutti quei siti che hanno scelto di impiegare le cosiddette tecniche "black hat" per ottenere un migliore posizionamento nei risultati di

46 DeMers 2013

47 <http://www.google.com/analytics/features/advertising.html>

48 Slegg, 2013

ricerca. Queste tecniche consistono tutte nel fare uso di metodi fraudolenti di cui si parlerà più approfonditamente nel paragrafo dedicato alla lotta continua da parte di Google allo spam. Secondo le stime riportate dal blog ufficiale di Google Search⁴⁹ l'aggiornamento di Google Penguin ha influenzato circa il 3.1% delle ricerche effettuate in lingua inglese, il 3% di quelle in tedesco, cinese e arabo mentre la percentuali di lingue più soggette a spam, come il polacco, raggiungono anche il 5%. Il 25 marzo 2012 l'algoritmo è stato nuovamente aggiornato con Penguin 1.1, obiettivo di questa modifica era sempre la penalizzazione dei siti Web facenti uso di tecniche manipolative per ottenere posizioni elevate nei risultati di ricerca.

I principali fattori analizzati dall'aggiornamento Penguin sono⁵⁰:

- La qualità dei *backlink* di un sito Web (link sponsorizzati, link presenti a fondo pagina, link dalle directory, link provenienti da *link exchange pages*, link provenienti da *network* di bassa qualità). Google riesce a determinare la credibilità e naturalezza di un link analizzandone la velocità di collegamento (data dal tasso di acquisizione di link)
- La provenienza dei *backlink*, soprattutto nel caso provengano da pagine Web di argomento estraneo a quello del sito in analisi
- Se troppi link indicano un sito Web con *anchor text* contenenti parole chiave perfettamente corrispondenti

Nel corso dell'anno Google ha rilasciato due aggiornamenti minori di Penguin seguiti poi, il 22 maggio del 2013, da uno di portata maggiore, denominato Penguin 2.0. Anche in questo caso i link presi di mira da Google erano quelli:

- provenienti da siti di bassa qualità
- situati su pagine web estranee al business del proprio sito Web
- link a pagamento
- link ricchi di parole chiave
- link dall'*anchor text* eccessivamente ottimizzato

PageRank

Ciò che differenzia essenzialmente Google da tutti gli altri motori di ricerca, è l'utilizzo di un algoritmo personalizzato, chiamato PageRank (da Larry Page, uno dei due fondatori dell'azienda). Lo scopo di questo meccanismo è quello di assegnare ad ogni pagina Web un punteggio di rilevanza. Questo punteggio dipende da vari fattori, tra cui:

- La frequenza e la posizione delle parole chiave all'interno della pagina Web: se esse appaiono solo una volta all'interno della pagina allora il sito riceverà un punteggio basso per

49 <http://insidesearch.blogspot.it/2012/04/another-step-to-reward-high-quality.html>

50 Zarokian 2012

quella *keyword*

- L'anzianità della pagina: nuovi siti Web vengono creati quotidianamente e non tutti durano più di qualche settimana. Google valuta maggiormente le pagine che hanno più storia alle loro spalle, in quanto più affidabili e con un pubblico già stabilito
- Il numero delle altre pagine Web che si collegano al sito in questione: Google analizza quante altre pagine si collegano ad un sito per capirne l'affidabilità e rilevanza

Dato che Google valuta positivamente i link presenti su un sito Web di collegamento ad un'altra pagina Web, è abbastanza difficile trovare il modo di aggirare il meccanismo. La maggioranza degli esperti SEO suggerisce a tale proposito di munire la propria pagine Web di ottimi contenuti, così facendo altre pagine Web sceglieranno di collegarsi alla stessa e, di conseguenza, si otterrà un buon punteggio nella pagina dei risultati di Google. Più un sito Web è oggetto di collegamenti da parte di altri siti, maggiore sarà il punteggio assegnato ad esso da PageRank. Allo stesso modo, i collegamenti attuati da siti con alti score dell' algoritmo consentiranno alle pagine oggetto del *link* di aumentare il proprio punteggio più velocemente. Sostanzialmente, PageRank è una sorta di "voto" che le pagine Web danno sull'importanza di un sito. Un link di collegamento conta come un voto di supporto, mentre l'assenza non conta come una votazione negativa, piuttosto come un'astensione⁵¹. Questi link di collegamento tra siti Web vengono chiamati *backlink* o *inbound link*. Il PR (PageRank) di un sito è visualizzabile o mediante la toolbar di Google⁵² o tramite diversi servizi disponibili online che si appoggiano ai database del motore di ricerca. Nel primo caso la toolbar mostra dei valori simbolici che possono andare dal minimo 0 (barra bianca) al massimo, 10 (barra verde). Nel caso in cui la barra si presenti di colore grigio allora può essere che la pagina non sia ancora presente nel database, o che ne sia stata eliminata, o che attualmente Google stia processando ed effettuando dei test⁵³.

2.2.3. La lotta allo spam

Google utilizza numerose tecniche per impedire ai webmasters di ingannare il sistema per ottenere un più alto posizionamento nelle SERP. Per esempio, se una pagina Web aggiunge numerosi link a più siti, il suo potere di voto diminuisce. Un sito Web dall'elevato PageRank con molti link in uscita può avere meno rilevanza di una pagina situatasi più in basso in classifica, con solo uno o due link in uscita. Nel momento in cui un webmaster si accinge a inserire il suo indirizzo URL su Google, verrà incoraggiato a verificare se il suo sito risponde alle norme di qualità impostate

51 <http://www.sirgroane.net/google-page-rank/>

52 Scaricabile a questo link: <http://toolbar.google.it>

53 Taverniti, 2011, pp. 60-61

dall'azienda. Queste norme sono le seguenti:

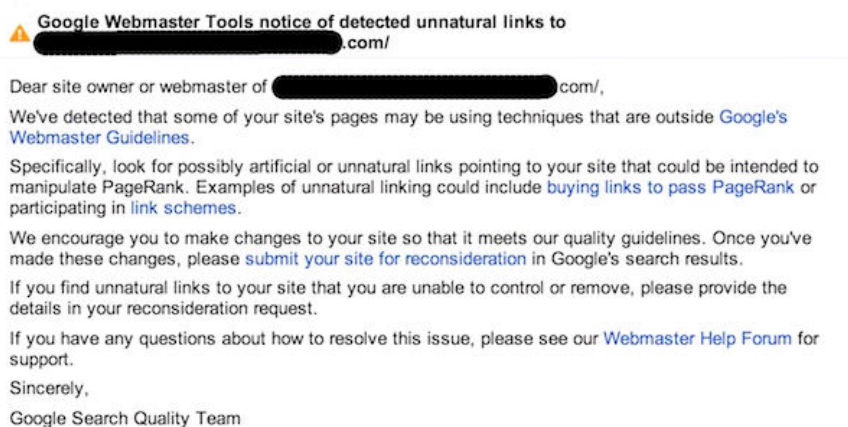
“Progetta le tue pagine per gli utenti, non per i motori di ricerca, non ingannare i tuoi utenti, evita i trucchi per migliorare il posizionamento nei motori di ricerca. Una buona regola generale è chiederti se ti sentiresti a tuo agio nello spiegare il tuo comportamento a un sito Web concorrente o a un dipendente di Google. Un altro test utile consiste nel chiederti se quanto stai facendo sarà d'aiuto ai tuoi utenti e se lo faresti ugualmente se i motori di ricerca non esistessero. Pensa a ciò che rende il tuo sito Web unico, prezioso o coinvolgente. Fai in modo che il tuo sito Web si distingua dagli altri nel tuo campo”⁵⁴

Le tecniche Black Hat

Si è già accennato di queste tecniche nel paragrafo dedicato alla spiegazione del funzionamento di Google Penguin; queste tecniche vengono sviluppate da alcuni webmaster con intenzioni di spamming. Le principali sono le seguenti:

- *Keyword stuffing*: con questa terminologia si intende letteralmente “l’imbottitura di parole chiave”
- Testo e contenuti generati automaticamente mediante strumenti come RSS o *frames*, inutili per l’utente
- Link non naturali, cioè *inbound* o *outbound* tra il sito Web in questione e altre pagine tramite strumenti automatici o tramite compravendite di link. Le tecniche di scambio link sono tollerate se effettuate in maniera equa e non eccessiva
- *Cloaking* e *sneaky redirect* (traduzione di mascherarsi e re-indirizzamento subdolo). Tramite l’utilizzo di queste metodologie il webmaster riesce a dare un contenuto diverso al suo sito Web a seconda che ad accedervi sia un *crawler* o un utente effettivo
- *Doorway pages*, vengono create dal webmaster per ottenere un ottimo ranking utilizzando specifiche parole chiave, ma in realtà il loro uso rimanda l’utente ad un sito diverso
- “Comportamenti maligni”, come scaricare virus, malware, o raccogliere *sensitive data* dagli utenti **IMMAGINE NOTIFICA GOOGLE**

54 <https://support.google.com/webmasters/answer/35769?hl=it#3>



2.3. I profitti di Google

A differenza di altre compagnie Internet, Google ha metodi diversi di generare dei profitti oltre agli investimenti privati o la vendita di quote del suo mercato. Esistono 3 modalità di partnership con venditori o pubblicitari: Google Checkout, Google AdWords e Google AdSense.

2.3.1. Google Checkout

Si tratta di un servizio creato per poter rendere gli acquisti online ancora più semplificati sia per il consumatore che per il venditore. I potenziali acquirenti possono creare un account completamente gratuito e durante la compilazione dei dati verrà richiesto di inserire il codice di una carta di credito o debito, che Google salverà in un database sicuro. Quando l'utente visita la pagina di un venditore iscritto al servizio, prima di terminare l'acquisto può scegliere l'opzione Google Checkout che faciliterà la transazione. Questo significa che l'utente non dovrà inserire il codice della propria carta ogni volta che effettua un acquisto. I venditori allo stesso modo possono creare un profilo su Google Checkout gratuitamente, anche se dall'agosto del 2008 è stata inserita una commissione del 2% assieme ad una tassa di 20 centesimi di dollaro per ogni transazione. Ad esempio se un cliente spende 10 \$ per un prodotto, Google riceverà 40 centesimi per quell'acquisto.

2.3.2. Google AdWords

La seconda modalità in grado di generare profitti è Google AdWords ed è più collegata all'interesse di questa tesi. Questo strumento consente ai pubblicitari di inserire le loro pubblicità su Google includendo una lista di parole chiave riguardanti il prodotto, il servizio o il business dell'azienda. Quando un utente effettuerà una ricerca utilizzando una o più di queste *keywords*, la pubblicità apparirà nella SERP sulla colonna di destra. Ogni qualvolta l'utente cliccherà sulla

pagina in questione, Google riceverà una somma dal pubblicitario.

2.3.4. Google AdSense

Si tratta di un servizio simile ad AdWord, anche se il webmaster può scegliere di mostrare la pubblicità direttamente sul suo sito invece che a margine nelle SERP. I *crawler* di Google scansionano il sito e analizzano il contenuto, dopodiché Google seleziona le pubblicità che contengano delle parole chiave rilevanti al sito del webmaster. Questi successivamente può modificare a piacimento il posizionamento e il colore della colonna laterale contenente le pubblicità. Tutte le volte che qualcuno clicca sulla pubblicità, il webmaster riceve una porzione dei profitti mentre Google ottiene il resto.

Tramite entrambi AdWords e AdSense, la strategia di Google è di fornire una pubblicità molto selezionata agli utenti. Il ragionamento seguito dall'azienda è che trovare e mostrare annunci pubblicitari rilevanti a quello che gli utenti stanno già cercando può aumentare le possibilità di click sul banner⁵⁵. Google ha sempre sostenuto di non avere alcuna intenzione di mettere in vendita gli spazi dei primi risultati di ricerca. Ogni pagina Web elencata nelle SERP di Google si trova in quella determinata posizione per via di diversi fattori, fra cui quello più determinante è costituito dal punteggio conseguito in seguito all'analisi svolta da PageRank. Gli altri spazi che vengono venduti dall'azienda tramite AdWords e AdSense, risultano evidenziati, con i banner incasellati in colori diversi, in modo da allertare gli utenti della differenza tra un sito sponsorizzato e uno non sponsorizzato.

The image shows a Google search results page for the query "vacanze a venezia". The search bar at the top shows the query and a search button. Below the search bar, there are navigation tabs for "Web", "Images", "Maps", "Shopping", "Videos", and "More". The search results are displayed in a grid format. On the left side, there are several organic search results, including "1375 Hotel Venezia da 33€ - trivago® Hotel Vacanze fino a -78%", "In Vacanza Spendendo Poco - Prova un Affitto Vacanze a Venezia", "Appartamenti Per Vacanze - Economici - In Centro a Venezia", "Vacanze Venezia, dove dormire e cosa visitare a Venezia", and "Vacanze Venezia - Viiaggi Venezia - recensioni e consigli - TripAdvisor". On the right side, there are several advertisements, including "Venice Hotels - save 70%", "Case Vacanza Venezia", and "Appartamenti a Venezia". The advertisements are highlighted with red boxes.

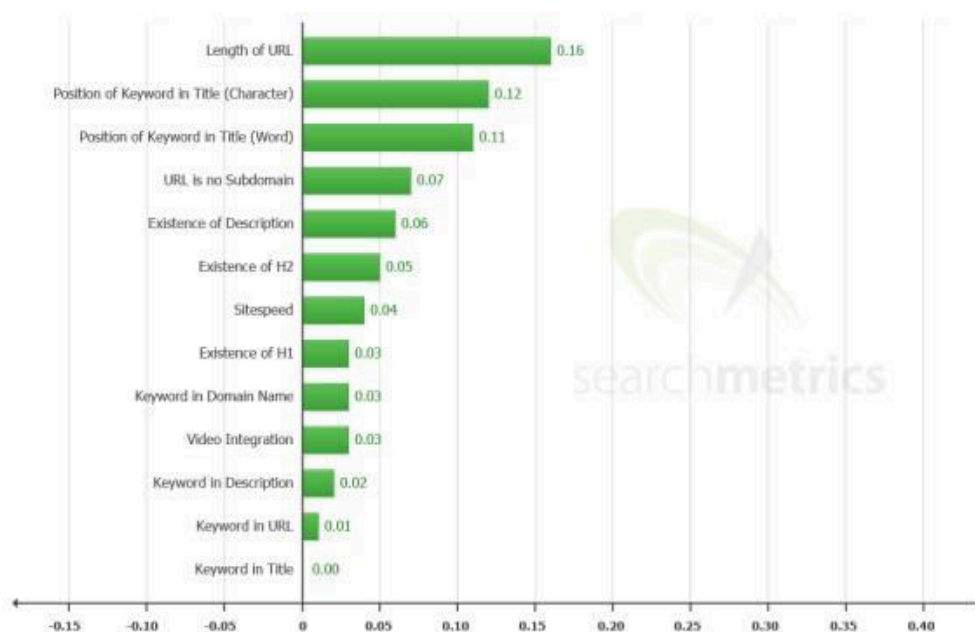
55 <http://www.google.com/intl/en/ads/>

2.4. Google Ranking Factors

Nel mese di luglio 2013 è stato pubblicato il tanto atteso studio annuale condotto da Searchmetrics⁵⁶ dal titolo “SEO Ranking Factors – Rank Correlation 2013”⁵⁷ con lo scopo di definire e valutare i fattori aventi alti coefficienti di correlazione di *rank*, onde rendere possibile un'analisi più approfondita degli algoritmi dei motori di ricerca. La raccolta dei dati e la loro valutazione fanno riferimento a Google.com nella prima metà del 2013 e in seguito nei grafici esplicativi vengono comparati a quelli del 2012 per individuare eventuali trend di sviluppo positivo o negativo. Appare evidente a chi scrive che la posizione e l'ordine dei risultati di ricerca non possono dipendere solo dalle parole chiave immesse nella finestra di ricerca, ma anche dalla storia personale di ricerca determinata dal comportamento dell'utente. Oltre a ciò avranno una certa influenza fattori come l'eventuale connessione tramite un profilo a Google o l'utilizzo di Google Chrome, la provenienza dell'IP e i dati memorizzati nei “cookie”⁵⁸.

I risultati sono stati divisi per macro-categorie che verranno fedelmente riportate qui di seguito.

2.4.1. On-page coding



56 Searchmetrics.com è una compagnia leader a livello mondiale nella fornitura di software finalizzati alla ricerca e analisi social per agenzie e aziende on-line. Il software utilizzato, Searchmetrics Suite TM, consente agli utenti di migliorare la loro visibilità, quote di mercato, ricavi e profitti applicando strategie ottimali nel marketing del mondo digitale

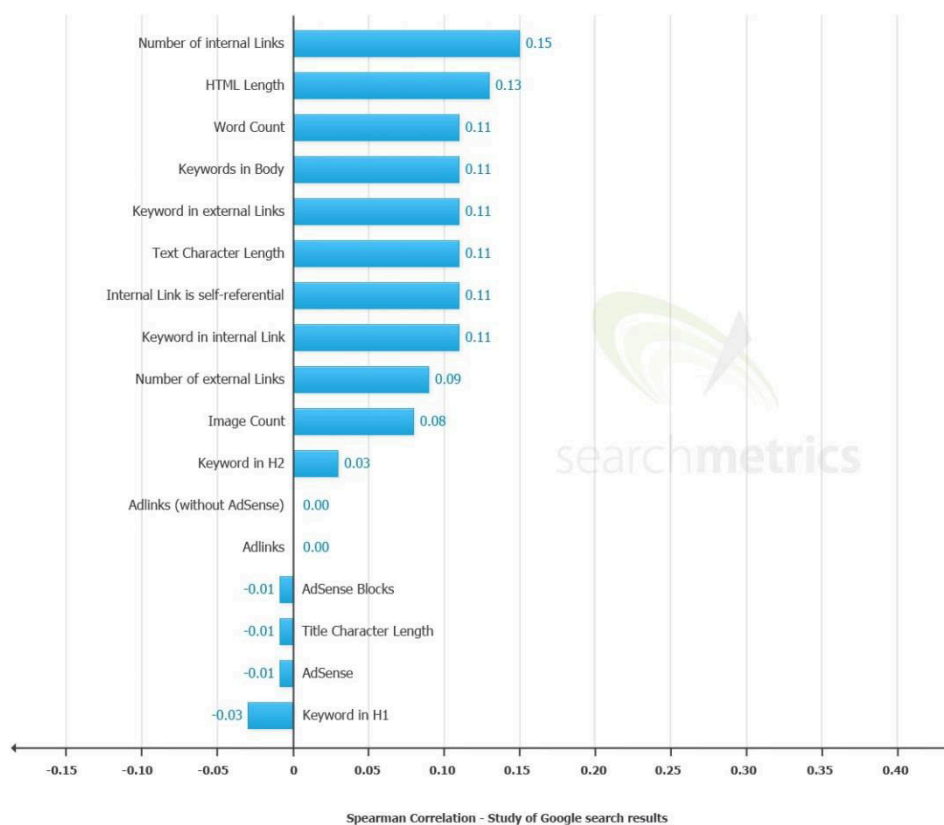
57 <http://www.searchmetrics.com/en/services/ranking-factors-2013/>

58 Citando la definizione fornita da Wikipedia: “in informatica i cookie HTTP (più comunemente denominati Web cookie, tracking cookie o semplicemente cookie) sono stringhe di testo usate per eseguire autenticazioni automatiche, come memorizzazioni di informazioni specifiche riguardanti l'accesso a siti web preferiti”

In questa sezione si fanno riferimento alle basi tecniche nella costruzione di una pagina Web, e i fattori che consentono di ottenere un buon ranking. È noto che l'attuazione di questi criteri non ha più un marcato effetto positivo sul ranking, ma il non attuarli porterebbe sicuramente risultati negativi. Dal grafico è possibile dedurre numerosi fattori, se si vuole un po' banali, ma comunque importanti ai fini di un buon posizionamento. Mentre fattori come la localizzazione delle *keywords* all'interno della pagina e del titolo hanno sempre maggiore importanza, altri come la presenza di parole chiave nell'URL o nel nome del dominio sono scesi agli ultimi posti. Secondo lo stesso studio ciò è avvenuto a causa dell'introduzione di due nuovi aggiornamenti di algoritmo da parte di Google nel 2012. Si parlerà a breve nello specifico dell'aggiornamento di Penguin, per ora ci si limita a dire che il suo principale scopo era la riduzione dello spam presente sul Web, come ad esempio siti internet con backlink di bassa qualità o domini con link provenienti dalle cosiddette "link farm", generatori automatici di collegamenti fittizi. Conseguentemente, c'è stato un ridimensionamento dei link composti unicamente da parole chiave nell'anchor text. Nel momento in cui i siti facenti uso di questi link di dubbia provenienza sono scalati nei punteggi finali, l'intero dominio è stato coinvolto, e ciò per via della presenza delle principali parole chiavi all'interno del suo title. Il secondo aggiornamento che pare aver influenzato negativamente lo sviluppo dei due fattori è stato rilasciato da Google nel settembre del 2012 e viene denominato EMD Update (Exact Match Domain). Questo aggiornamento non è stato ancora applicato fuori dagli Stati Uniti⁵⁹ e, a differenza dell'update di Penguin, basato principalmente su link e parole chiave, analizzando off-page la qualità dei backlink, questo algoritmo oltre ad agire sulle keywords, viene implementato direttamente on-page e si concentra sui domini. Altri fattori di ranking importanti sottolineati dallo studio sono la localizzazione delle parole chiave nel titolo, che devono essere posizionate il più vicino possibile ad inizio pagina, ed infine la velocità del sito Web che deve caricarsi rapidamente.

2.4.2. On-page content

⁵⁹ L'EMD update non è stato implementato in Europa probabilmente perché la maggioranza dei domini sospetti era già stata segnalata e fatta regredire nei risultati grazie all'utilizzo dell'aggiornamento di Penguin, rendendo superflui ulteriori aggiornamenti nel motore di ricerca

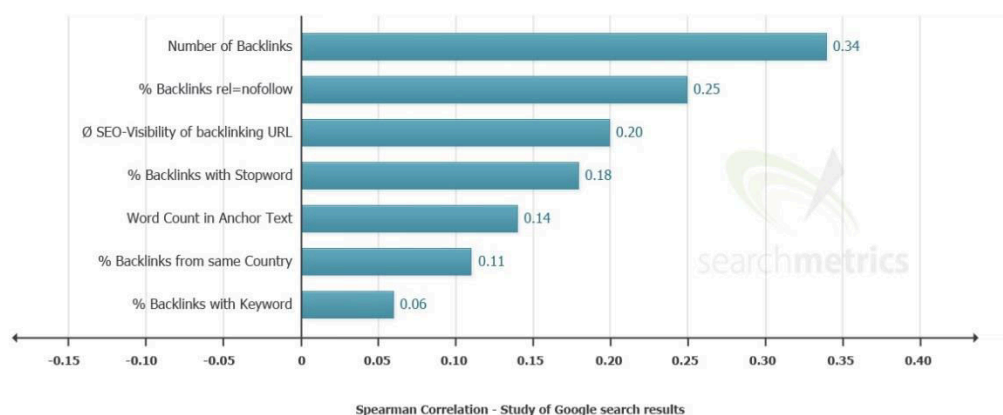


Il motore di ricerca sarà automaticamente in grado di analizzare e categorizzare con grande facilità i dettagli tecnici delle pagine Web, mentre per quanto riguarda i contenuti la situazione si fa più complicata. Le caratteristiche dei contenuti sono spesso molto complesse e la loro valutazione si muove all'interno di una scala rispetto alla quale Google deve dare una valutazione di pertinenza. Dopotutto è compito di un motore di ricerca valutare e classificare una pagina Web in base alla rilevanza dei contenuti piuttosto che in base all'implementazione tecnica di regole.

I principali fattori racchiusi nella classe "on-page content" sono tre e sono riconducibili a tre classi diverse, ovvero la qualità dei contenuti, i collegamenti e la pubblicità. Il primo è "image count", difatti lo studio di Searchmetrics.com ha dimostrato che l'utilizzo di immagini e video all'interno di una pagina Web la rende più attraente non solo agli occhi dell'utente, ma anche a quelli di Google. Più immagini e contenuti mediali vengono utilizzati da un Web site, migliore sarà il suo ranking. Altro fattore su cui ci si vuole soffermare è quello relativo ai collegamenti interni delle pagine Web. I siti internet che godono del migliore *ranking* hanno una struttura interna di link nettamente superiore di quella dei siti posizionatisi successivamente. La media dei link presenti in

una pagina aumenterà nel caso di siti web dotati di indici e strutture di link in fondo alla pagina. È importante ai fini di un buon *ranking* non solo il numero dei link ma anche le parole chiave che portano a collegamenti su pagine diverse del medesimo sito. È una regola normalmente seguita dai webmasters quella di utilizzare i cosiddetti “hard link” all’interno del sito (con l’utilizzo di parole chiave) mentre “soft link” vengono utilizzati verso l’esterno (utilizzando più parole generiche nell’anchor text, brand links, ecc.). Nell’apertura di questo paragrafo si è detto di come i dati raccolti dallo studio vengano poi messi in correlazioni con le medesime voci relative all’anno 2012. Ebbene il terzo fattore riguarda l’utilizzo di pubblicità e nel corso del 2012 si erano registrate correlazioni chiaramente negative tra *advertising* (anche utilizzando AdSense) e l’ottenimento di un buon *ranking*. Questo trend è andato significativamente a modificarsi nel corso del 2013; attualmente i siti posizionatisi ottimamente nella SERP di Google hanno integrato al loro interno l’utilizzo di pubblicità.

2.4.3. Backlinks

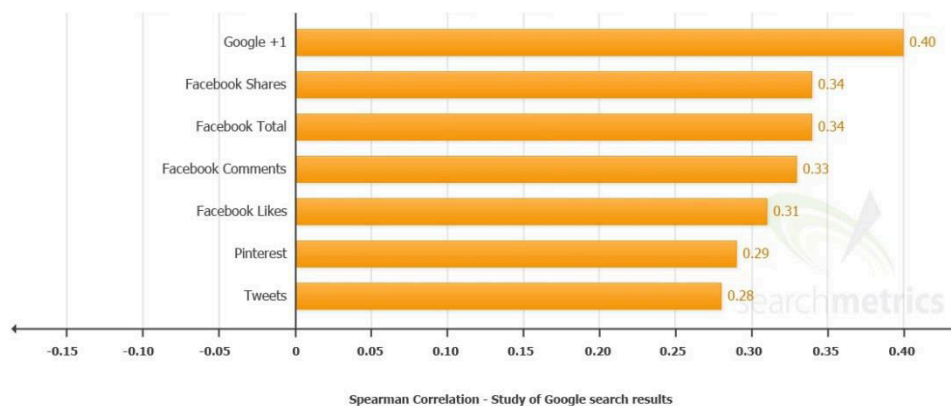


Si è già detto abbondantemente dell’importanza della qualità dei *backlink*, soprattutto visto quanto Google faccia affidamento sull’algoritmo *PageRank* nella determinazione del ranking delle pagine Web. Il posizionamento di un sito Web è determinato non solo dal numero dei *backlink*, ma anche dalla qualità della struttura del link che è determinata da vari fattori. Dallo studio è emerso che i siti posizionatisi meglio nel corso del 2013 hanno un elevato numero di *backlinks* e, in particolare, link di collegamento provenienti da siti Web dotati di un’alta visibilità SEO⁶⁰ hanno molto più

60 Lo studio ha sottolineato che websites classificatisi al primo posto per determinate keywords tendono ad aver link da URL con una visibilità media di SEO di oltre 65000, includendo collegamenti a Wikipedia, mentre quelli

valore di *backlink* provenienti da siti non adeguatamente potenziati a livello di applicazioni di tecniche SEO. Per questi motivi i *backlinks* continuano ad essere uno dei fattori più critici delle metriche di analisi SEO. Attualmente risultato importanti non tanto il numero dei link, ma la qualità. Google ha recentemente rivalutato numerosi fattori riguardanti le parole chiave, soprattutto in seguito alle ultime implementazioni di Penguin. Come conseguenza siti ben posizionati utilizzano un'unione molto equilibrata di "nofollow" links, anchor texts dotati di più parole, link text con condivisione tramite stop word e keyword links. Anche la qualità della provenienza dei link si è già detto essere sotto attenta osservazione, gli algoritmi di Google sono sempre più adepti all'identificazione delle *link farms*.

2.4.4. I social network



L'utilizzo dei social network ha un elevatissimo livello di correlazione con un buon *ranking*, ciò provoca poche sorprese in quanto già largamente evidente dallo studio condotto nel 2012. In particolare l'impiego di Facebook e Twitter mostrano un livello di correlazione superiore a quello dell'anno precedente. La nuova aggiunta del network di Google, Google+, mostra elevatissimi risultati per le pagine Web che hanno scelto di farne uso. Nello studio del 2012 il nuovo social network non era stato nominato in quanto non erano stati ancora raccolti sufficienti dati, eppure nell'anno corrente 2013 pare che le segnalazioni di Google+ siano appena seconde ai *tweets* negli Stati Uniti. Contemporaneamente anche l'impiego della photo community Pinterest ha portato ottimi risultati. Tra l'altro pare anche interessante riportare che nello studio viene effettuata una previsione documentata e giustificata secondo cui, se il tasso di crescita delle condivisioni studiato

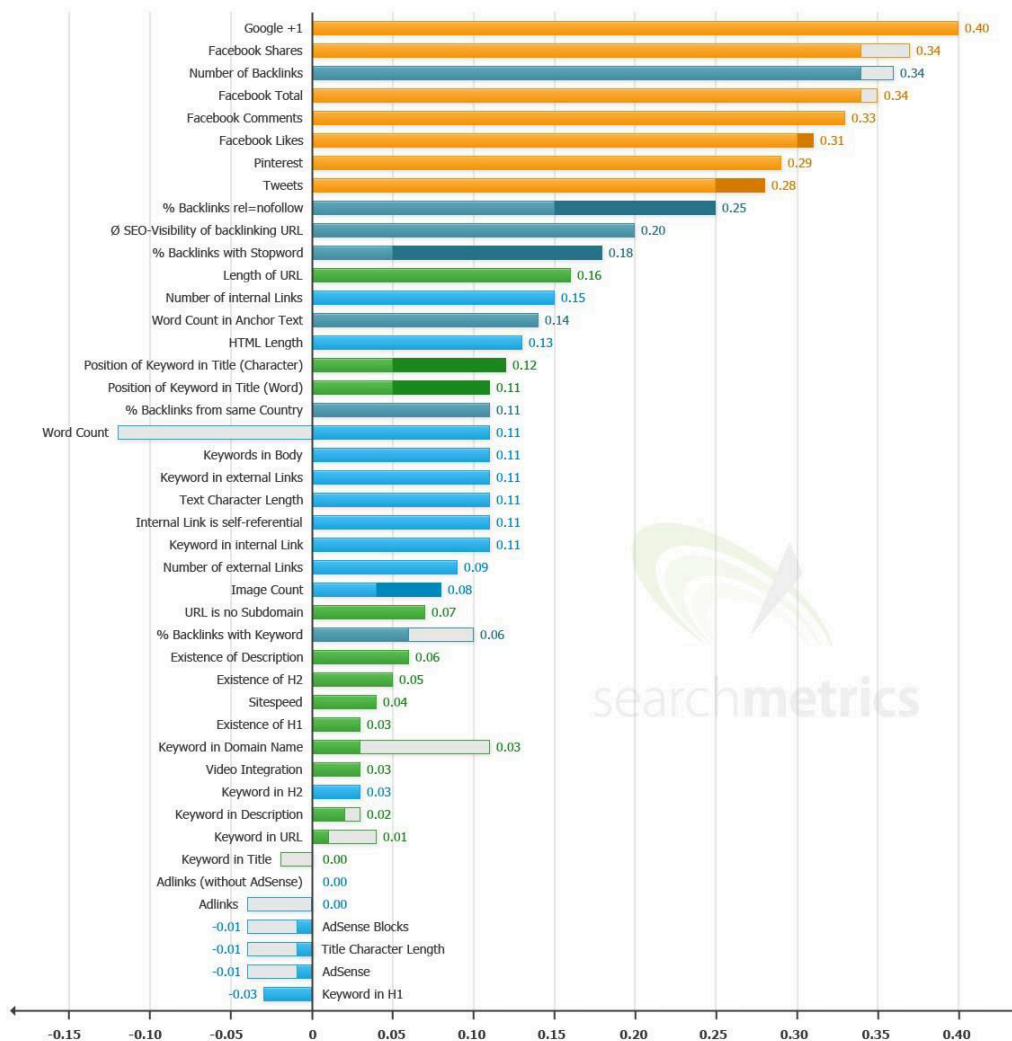
finora si mantiene costante, nel febbraio del 2016 Google+ supererà Facebook.

2.4.5. Il “brand factor” e il caso particolare Wikipedia

Dal punto di vista del motore di ricerca Google, le grandi marche svolgono un ruolo speciale e vengono fortemente preferite nelle SERP. Per i siti non soggetti a nessun *branding* classificarsi ai primi posti è estremamente faticoso e necessita di adempiere a numerosi requisiti; il cosiddetto “brand factor” fa sì che in cima ai risultati figurino sempre determinati siti indipendentemente dalla loro implementazione dei fattori già visti. Questo perché la loro gestione dei backlink è migliore, hanno un’ottima visibilità SEO, ma soprattutto i brand ricevono molti segnali social.

Wikipedia ha un impatto enorme sul ranking dei siti ad essa collegata, basti pensare che qualsiasi sia la ricerca condotta da un utente, quasi sicuramente tra i primi due risultati di ricerca apparirà un suo link, assieme al brand link.

Come sintesi finale, a conclusione del discorso, si è scelto di riportare l’immagine contenente tutti i principali Google Ranking Factors di cui si è già effettuata la descrizione con l’indice di correlazione rispettivo.



Spearman Correlation - Study of Google search results

■ Social
 ■ Backlinks
 ■ Onpage (technical)
 ■ Onpage (content)

ANALISI DELLE TECNICHE SEO

Le tecniche SEO (Search Engine Optimization) consentono di aumentare la visibilità di un sito internet tramite un migliore posizionamento nei risultati forniti dai motori di ricerca. È evidente che se una pagina Web si classifica ai primi posti nei risultati di ricerca godrà di maggiori visualizzazioni e riceverà quotidianamente un elevato numero di visitatori. L'applicazione delle tecniche SEO rientra in un ambito più ampio detto SEM, *Search Engine Marketing*, e, in quanto strategia di *web marketing*, la SEO considera e analizza il funzionamento dei motori di ricerca. In questo campo è di fondamentale importanza sapere determinate informazioni, come che cosa viene esattamente cercato dai visitatori, i termini di ricerca o parole chiave inseriti nei motori di ricerca e anche l'eventuale preferenza di determinati *search engine*. L'ottimizzazione di un sito Web può avvenire tramite varie modalità, si possono apportare modifiche ai contenuti tramite codifiche HTML, aumentare la rilevanza della pagina mediante specifiche parole chiave e rimuovere gli ostacoli alle attività di indicizzazione dei motori di ricerca. Un'altra tattica SEO riguarda la promozione di un sito per aumentarne il numero di *backlinks*, o link in entrata. Tutte queste modalità verranno analizzate nel dettaglio in questo capitolo, ma prima si vuole spiegare il perché pare necessario attualmente interessarsi e approfondire l'argomento.

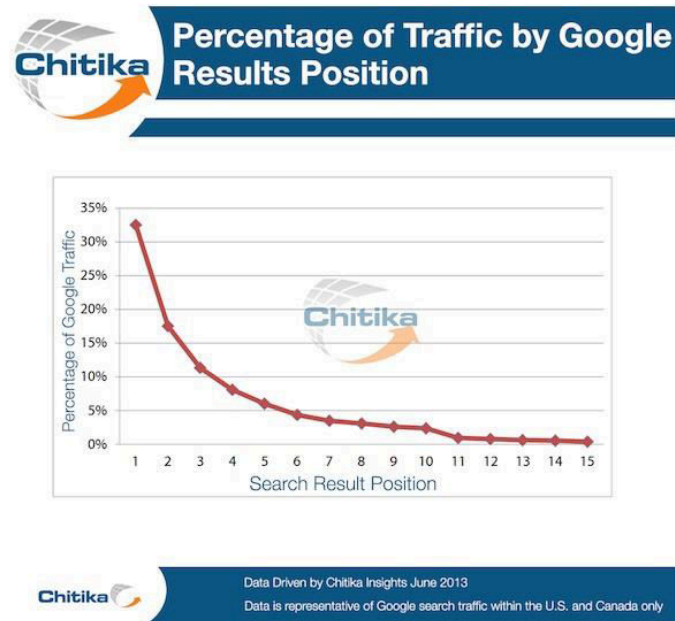
1. L'importanza delle tecniche SEO

Le tecniche SEO vengono utilizzate in vari ambiti, sia di tipo accademico, che di profitto. Nel primo caso, è di notevole importanza per i ricercatori assicurarsi che i loro articoli vengano indicizzati dai motori di ricerca (Google Scholar, SciPlore.org, PubMed,..) in modo che i loro studi siano più facilmente fruibili dalla comunità. Oltre all'indicizzazione, come si vedrà in seguito, è fondamentale interessarsi anche alla posizione in cui gli articoli vengono visualizzati nell'elenco dei risultati; difatti come in ogni classifica di risultati di ricerca, gli articoli posizionatisi ai primi posti avranno maggiori probabilità di essere letti⁶¹. A livello di profitto, rientra nei migliori interessi delle aziende avere un elevato *ranking* nei risultati di ricerca di Google, consentendo una maggiore visibilità e conseguentemente un elevato numero di visitatori, potenziali clienti. Numerosi studi (come quello svolto dal network di pubblicità online Chitika⁶²) testimoniano l'enorme quotazione ottenuta dalle pagine Web posizionatisi in alto nei risultati di ricerca organici di Google. In questo studio pubblicato nel giugno del 2013 si legge che il sito web classificatosi primo in una determinata ricerca, riceve circa il 33% del traffico, quasi il doppio del

61 J. Beel, B. Gipp, E. Wilde, 2000, p. 1

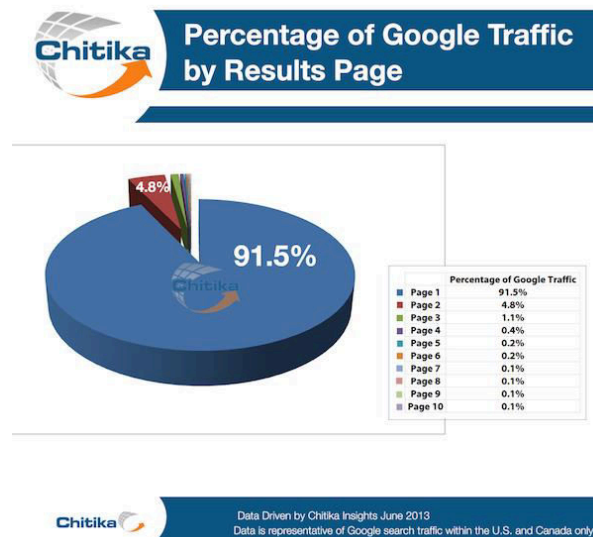
62 <http://chitika.com/google-positioning-value>

secondo classificato, che riceve “solo” il 18% del secondo.



Uno studio simile, svolto sempre dal team Chitika nel 2010, aveva già anticipato quanto oggi è ormai riconosciuto da tutti gli esperti, convalidando l'importanza delle tecniche SEO per le imprese dotate di un sito Web. È evidente che la pagina Web posizionatasi per prima avrà un notevole vantaggio rispetto alle ultime classificate e questo trend pare si manterrà anche per i tempi a venire, visto come già lo studio del 2010 abbia profetizzato i risultati odierni.

Allo stesso modo i siti Web posizionatisi nella prima pagina dei risultati di ricerca godranno del 92% di tutto il traffico di ricerca, mentre la seconda pagina avrà il 4.8%, e via via si scala fino allo 0.1% delle ultime pagine.



È anche vero che non si può considerare il *ranking* come unico indicatore di successo SEO, in quanto le classifiche dei risultati di ricerca possono variare (all'interno di una precisa *query*) da un utente all'altro; i dati dello studio dimostrano come, ad ogni modo, le prime posizioni ottengano la maggioranza della quota di traffico. Si cita a questo proposito uno studio svolto nel 2012 da Compete.com, azienda leader nelle consulenze di Digital Marketing, che mostra come il 53% degli utenti clicchi direttamente il primo risultato di ricerca ottenuto, spesso senza nemmeno guardare gli altri URL forniti⁶³.

2. Le principali tecniche di Search Engine Optimization

Nei prossimi paragrafi si vogliono illustrare e spiegare al meglio le principali tecniche volte al potenziamento del ranking di un sito Web. Prima di cominciare vale la pena ricordare che, qualsiasi tecnica si scelga di implementare, non si può essere in grado di valutarne l'efficacia a meno che non si decida di monitorare costantemente il proprio page rank, prima durante e dopo l'applicazione di queste tecniche. A questo proposito, si raccomanda l'utilizzo di strumenti quali Alexa⁶⁴ o Google Toolbar⁶⁵, il cui compito è quello di svolgere proprio questo tipo di analisi⁶⁶. In particolare, in seguito all'implementazione delle tecniche di ottimizzazione on-page, si consiglia di monitorarne l'andamento sfruttando questi due strumenti disponibili online:

- SEO Automatic⁶⁷, è uno strumento gratuito nella sua versione base che consente, in seguito all'inserimento dell'URL della pagina di cui si vuole compiere l'analisi, di monitorarne l'andamento SEO;
- SEO Doctor⁶⁸, un plug-in per Mozilla Firefox, che fornisce immediate feedback visivi e permette di analizzare l'andamento dei principali fattori SEO

Oltre a questi tools, per un webmaster è possibile consultare i file di log del web server (Apache, IIS...), che sarebbero dei file di testo composti da un elenco degli accessi di un sito. Tra le varie informazioni a disposizione è possibile recuperare, tramite la stringa referrer, la pagine web da dove è partita la richiesta dell'utente e anche le keyword digitate per raggiungere quel sito. Tramite questo strumento è possibile valutare le prestazioni della pagina rispetto a determinate parole chiave⁶⁹.

Tornando alle tecniche SEO, si è pensato di semplificarne la presentazione dividendole in tre

63 <http://searchenginewatch.com/article/2215868/53-of-Organic-Search-Clicks-Go-to-First-Link-Study>

64 www.alexa.com

65 <http://www.google.com/toolbar/ie/index.html>

66 <http://www.businessinsider.com/10-basic-seo-tips-everyone-should-know-2010-1?op=1>

67 <http://www.seoautomatic.com/unique-tools/instant-seo-review>

68 <http://www.prelovac.com/vladimir/browser-addons>

69 Vasta, 2009, p. 5

grandi categorie:

- analisi di mercato
- analisi interna
- analisi esterna

2.1. Analisi di mercato

Questa analisi viene generalmente svolta per prima, in quanto risulta fondamentale analizzare il mercato sia per rintracciare eventuali competitors (virtualmente senza confini geografici) che per individuare i propri utenti. Ciò è reso possibile dall'utilizzo degli strumenti di Web Analytics. Questi strumenti consentono di conoscere a fondo e ottimizzare l'utilizzo del Web tramite analisi, metriche qualitative e statistiche di dati internet. Tra questi sistemi si ricordano Google Analytics, ShinyStat Free, Yahoo! Web Analytics e Weboscope Free⁷⁰.

Il posizionamento di un sito Web sul mercato online dipende quasi interamente dalle *keywords* digitate dagli utenti.

2.1.1. Utilizzo delle parole chiave e il keyword sniping

La comprensione dello strumento delle parole chiave è fondamentale in quanto è la prima modalità con cui gli utenti possono venire a contatto con i siti Web, confrontare varie aziende e successivamente effettuare una scelta di preferenza. Prima di approfondire il discorso sulle tipologie di parole chiave e tecniche varie di ranking, si è pensato di illustrare un po' schematicamente quali sono le fasi, e le problematiche, che un esperto SEO si trova a dover affrontare per poter applicare questo strumento. Il primo passo che viene effettuato riguarda la keyword research, in modo da poter capire quali possano essere i termini su cui valga la pena investire, in quanto più interessanti e suscettibili di ricerca da parte degli utenti. Questa azione viene attuata tramite l'utilizzo di strumenti quali Google Trends, Ubersuggest.org, Wordstream.com, ecc.; tramite questi tool si può capire quanta domanda ci sia effettivamente per ogni singola parola chiave. All'interno dello strumento Google Trends è possibile selezionare la voce "google search volume"; qui si possono selezionare e modificare vari parametri (quali il fattore tempo, lingua, device e location) e così facendo si ottiene una sorta di lista ordinata per rilevanza. Per ogni parola è necessario poi valutare il fattore competizione, se infatti una keyword ha un elevato google search volume, è comprensibile che essa appaia come più attraente agli occhi dei competitors e quindi sarà quella dai costi maggiori e su cui ci sarà più da scontrarsi all'interno

70 Vasta, 2009, pp. 34-41

del proprio settore per il clic dell'utente. È utile considerare anche il fattore variazione, infatti l'utilizzo di parole chiave quali "museo Rimini" non è la stessa cosa di "musei a Rimini" o "il museo a Rimini". Una volta realizzata una lista di parole chiave in ordine di priorità è utile inserire il discorso su un piano temporale. L'utilizzo di keywords quali "vacanze natalizie a Rimini" sarebbe ottimo durante i periodi invernali, o quelli immediatamente precedenti, ma in prossimità della stagione estiva è decisamente improbabile che qualcuno possa effettuare ricerche di questo tipo. È quindi essenziale svolgere un lavoro di ricerca di parole chiave continuativo nel tempo, magari ruotandole tra le diverse stagioni. Ultimo punto, non meno importante, sono le considerazioni legate all'uso che eventuali competitors fanno delle parole chiave; quindi quali utilizzano, come e quando lo fanno.

Le keyword possono essere generalmente raggruppate in due categorie:

- *Keyword long tail*, costituite da tre o più termini, ad esempio: "Università Ca'Foscari Venezia"
- *Keyword short tail*, costituite da uno o due termini, ad esempio: "Università Venezia"

Queste ultime sono sempre state quelle più utilizzate, data la loro generalità e semplicità, ma conseguentemente sono spesso le più inflazionate e di dominio dei siti con un posizionamento consolidato nella SERP, ovvero i siti più conosciuti. Un utente tende ad utilizzare *keyword long tail* quando è alla ricerca di qualcosa di ben definito, oppure quando una prima ricerca non ha prodotto i risultati sperati e quindi si cerca di raffinare gli URL ottenuti; per questi motivi il volume di traffico da essere veicolato è assai inferiore⁷¹. Esiste anche un sottoinsieme di *keyword long* in grado di veicolare una buona quantità di traffico, con un buon potenziale di redditività; si tratta delle cosiddette *keyword money tail*⁷². Esse sono oggetto di minore competizione rispetto alle *short tail* e garantiscono un alto tasso di conversione, ovvero consentono all'utente di arrivare con una ragionevole certezza alla pagina del prodotto o del servizio che sta cercando. Le parole chiave costituite da tre termini sono preferibili per i siti emergenti, in quanto i motori di ricerca premiano la specificità e un nuovo sito ottimizzato su *keyword long* può comparire tra i primi risultati, entrando effettivamente in competizione con siti più affermati e già conosciuti. L'importanza delle keyword risiede nella loro capacità di descrivere con specificità l'attività di cui si occupa l'organizzazione; per questo motivo è consigliabile scegliere quelle che meglio identificano la *core activity*, piuttosto che quelle più popolari e diffuse.

Il *keyword sniping* è una tecnica SEO mirata proprio all'ottenimento di un buon ranking tra i

71 Maltraversi, 2012, pag.105

72 L. Toscano, 2009, pp. 20-21

risultati di una ricerca mediante l'utilizzo di alcune keyword che rappresentano il giusto equilibrio tra specificità e genericità.

Trovare le parole chiave più adatte non è cosa semplice, per questo motivo è possibile aiutarsi tramite numerosi strumenti. Sceglierne alcuni tra i più conosciuti si segnalano:

- Google AdWord, strumento di cui si è già brevemente parlato nel primo capitolo a proposito delle modalità di profitto di Google. Lo strumento consente di determinare, per ogni keyword selezionata, il volume approssimativo di ricerche nell'ultimo mese o di effettuare una media calcolata sulle ricerche eseguite nell'ultimo anno. È possibile filtrare le misurazioni utilizzando vari parametri, come paese, lingua, ecc. e ciò dimostra come questo tool sia sufficientemente valido per suggerire le parole chiave correlate alla rispettiva popolarità;
- Google Trends, questo strumento, gratuito nella sua versione base, consente di analizzare gli andamenti di ricerca sia in senso temporale (volumi di ricerca nel tempo), sia spaziale (da quale paese è stata ricercata una determinata keyword) e consente di fare paragoni nella ricerca. Di questo strumento si farà ampiamente uso nel prossimo capitolo, quando si andranno ad analizzare le parole chiave del sito del Museo della Città di Rimini;
- Google Instant è uno strumento online che permette un miglioramento della ricerca in tempo reale. Nel momento in cui l'utente digita una parola chiave nella finestra di ricerca, automaticamente compaiono una serie di keyword che permettono di ottenere informazioni molto più velocemente di prima, poiché non si deve nemmeno finire di digitare quanto si sta cercando o premere "ricerca". Contemporaneamente la possibilità di vedere i risultati durante la digitazione consente di ottenere un feedback immediato sui termini di ricerca che si stanno impiegando⁷³.

Questi mezzi tecnologici possono essere utili nel suggerimento di nuove keyword e possono supportare (mediante dati numerici e percentuali) le varie opinioni nel momento della scelta, ma non dovrebbero andare a sostituirsi a sondaggi effettuati presso gli utenti stessi.

Chi meglio di un visitatore del sito Web sarebbe in grado di identificare esattamente:

- Cosa digitare per trovare una determinata società/organizzazione
- Cosa digitare su Google per ottenere determinati servizi
- Cosa inserire come parola chiave per cercare consigli o informazioni su eventi
- Quali siano le informazioni più importanti ottenibili dal sito Internet di un'azienda

Queste sono solo alcune delle domande a cui gli utenti potrebbero dare risposta utilizzando

73 <http://www.google.com/insidesearch/features/instant/about.html>

programmi come 4Q⁷⁴ o SurveyMonkey, che, come recita la pagina Web, si presenta come: “[...] lo strumento di indagine on-line più popolare al mondo. Inviare indagini gratuite sotto forma di sondaggi, questionari, feedback dei clienti e ricerche di mercato non è mai stato così facile”⁷⁵. Nel momento in cui si inviano i questionari alle *mailing lists* di aziende o musei, è possibile ottenere risposte da collaboratori, professionisti del settore, iscritti alla newsletter, ma anche potenziali visitatori da attirare, tramite una buona campagna SEM. Oltre ai questionari è possibile effettuare user testing, come focus groups o tramite strumenti online come Usability Hub⁷⁶ per verificare l’efficacia del proprio sito Web. Questa ricerca effettuata presso gli utenti può suggerire cosa le persone stiano cercando e contribuire all’identificazione del possibile visitatore, ma è necessario integrarla con gli strumenti di Web Analytics per capire il traffico relativo ad ogni parola ed eventuali difficoltà di utilizzo. L’identificazione dei *competitors* è resa possibile da alcuni *tools* forniti da software come SeoMoz⁷⁷ e Searchmetric⁷⁸, ma principalmente gli esperti SEO effettuano dei *benchmark* di riferimento con le *keyword* per il quale si intende competere.

2.1.2. Benchmarking

Quella del benchmarking costituisce una pratica molto diffusa negli ambienti lavorativi e porta alla misurazione dei risultati aziendali o di un determinato business, in relazione ai concorrenti. Sostanzialmente, si procede tramite confronto di ambiti molto simili e tramite l’utilizzo di sistemi di Web Analytics che offrono funzioni di benchmarking. Ad esempio, utilizzando la sezione *Analisi Comparativa* in Google Analytics, il sito Web viene messo a confronto con dei competitors di simili dimensioni e poi si verifica per ogni categoria il livello del proprio website. L’obiettivo è quello di individuare i punti di forza e debolezza dei concorrenti, analizzando alcuni fattori quali domain authority, età del sito, numero di pagine che lo compongono, visite, frequenze di rimbalzo, visualizzazioni della pagina, tempo medio sul sito, ecc.

In alternativa ShinyStatBusiness⁷⁹ consente di utilizzare l’opzione di geolocalizzazione, dove è più semplice visualizzare per ogni regione chi sono i propri competitori diretti. Il rovescio della medaglia è che però non è possibile selezionare un periodo predefinito di paragone, il confronto è in termini assoluti⁸⁰.

74 <http://www.4qsurvey.com>

75 <http://it.surveymonkey.com>

76 <https://usabilityhub.com>

77 <http://moz.com>

78 <https://www.searchmetrics.com>

79 http://www.shinystat.com/it/biz/info_biz.html

80 Vasta, 2009, pp. 185-186

2.2. Analisi interna (on-page optimization)

Questo tipo di analisi viene svolta dai Webmaster per fare in modo che la struttura complessiva di server, pagina e contenuto, permetta l'indicizzazione del sito. Ciò che in particolare si deve fare è eliminare ogni forma di ostacolo ad una maggiore visualizzazione. Come si è visto nel primo capitolo, è necessario che il sito sia accessibile e rintracciabile dai *crawler* e che il contenuto sia chiaramente leggibile. Questo tipo di analisi è molto "personale", nel senso che ogni sito presenta una situazione diversa, in base alle esigenze del settore all'interno del quale si opera.

Oltre a questo si vuole fare presente anche la nuova direzione di sviluppo riguardante la capacità di effettuare una correlazione semantica dei contenuti (e non solo tramite le *keywords*) negli algoritmi di indicizzazione. Questi studi riguardano un ambito informatico molto studiato che è quello dell'*Information Retrieval*, ma se ne intende parlare più approfonditamente nell'ultimo capitolo che riguarda gli orizzonti futuri di sviluppo delle tecniche SEO.

2.2.1. Il nome del dominio

Si tratta, assieme alla scelta del provider, di una decisione molto importante, in quanto costituisce una proiezione ottimale e focalizzata dell'argomento di cui tratta il sito Web. Il nome del dominio è costituito dal nome dell'host e dal TLD (Top Level Domain). Si vuole fare un semplice esempio chiarificatore, il sito Web dell'università Ca'Foscari è:

www.unive.it

In questo caso, il nome dell'host sarà UNIVE, mentre “.it” costituisce il TLD. Questo dominio di livello superiore rappresenta o un'area geografica (.it Italia, .de Germania,..) oppure una categoria (.com commerciale, .org organizzazioni non commerciali,..)⁸¹.

Il nome del dominio deve essere breve, facile da ricordare e identificativo sia del nome dell'azienda/organizzazione per cui si sta lavorando, sia della tipologia di attività. Viene suggerito da alcuni testi⁸² di registrare lo stesso dominio con estensioni diverse (come www.unive.it, www.unive.com, www.unive.org, ..) per evitare che qualche altro webmaster possa appropriarsi del medesimo nome⁸³. Generalmente viene consigliato di effettuare la scelta tenendo conto dell'effetto comunicativo agli utenti e al target medio cui ci si rivolge. È possibile anche aggiungere dettagli come la località di appartenenza (se rilevante) come nel caso dell'università,

81 Maltraversi, 2012, p. 79

82 Maltraversi, 2012 p. 84

83 Esiste proprio una tecnica chiamata cybersquatting che consiste nell'appropriarsi di nomi di domini famosi, appartenenti a star o aziende di successo, per poi rivenderli a prezzi altissimi

l'importante è che sia un nome facilmente pronunciabile e comprensibile⁸⁴. Esistono vari tool a disposizione del Webmaster nel momento della scelta del dominio, i migliori che vengono riportati in letteratura⁸⁵ sono:

- DomainFellow.com che a partire da una keyword permette di aggiungere prefissi e suffissi;
- MakeWords.com consente di determinare la lunghezza e lingua di un dominio oltre a prefissi e suffissi;

Mentre per poter verificare eventuali anomalie o punti critici in seguito a una scelta erranea:

- Intodns.com
- Robtex.com

Come si è visto nel paragrafo dedicato ai fattori di posizionamento su Google, la scelta di “hard keyword” (ovvero di parole chiave molto forzate) può contribuire a posizionare meglio temporaneamente un sito Web, ma ci saranno sempre fluttuazioni e nei settori più competitivi non si garantiscono risultati di grande importanza. Google ha pubblicato nel 2003 un paper nominato [Systems and methods for detecting commercial queries](https://www.google.com/patents/US8046350)⁸⁶ all'interno del quale si legge che sono preferiti nomi di dominio senza trattini, con al massimo due separatori “-” e contenenti query non commerciali.

2.2.2. Keyword density e il rischio di keyword stuffing

La KD è data dal rapporto tra il numero di volte in cui una determinata parola compare nella pagina e il numero totale delle parole presenti all'interno della pagina, espressa in percentuale. Questa tecnica nasce dall'idea per cui ripetendo più volte una parola chiave all'interno di un documento si possano ottenere migliori posizioni per quella determinata parola. A livello teorico questo ragionamento potrebbe funzionare, ma bisogna fare molta attenzione a non esagerare con le parole ripetute, onde evitare il fenomeno di *keyword stuffing*. Questo fenomeno si verifica quando un webmaster perde di vista la priorità delle parole chiave più importanti e conseguentemente il sito viene eliminato dai risultati di ricerca, in quanto Google (o gli altri *search engine*) lo percepiscono come spam. Una pagina Web è viene trattata come spam nel caso in cui cerchi di classificarsi ai primi posti per più categorie, spesso di argomento non correlato, oppure ripeta eccessivamente una keyword⁸⁷.

2.2.3. Meta Tag

84 Taverniti, 2011, pp. 22-23

85 Maltraversi, 2012, pp. 88-94

86 <https://www.google.com/patents/US8046350>

87 Maltraversi, 2012, pp. 178-179

I meta tag costituiscono solo una parte del grande puzzle algoritmico che i principali motori di ricerca guardano quando si deve decidere la rilevanza dei risultati da mostrare agli utenti che hanno digitato una *query*. I meta tag sono porzioni di codici di pagina situati tra i tag posti ai due limiti di apertura e chiusura del codice HTML di un documento. Il testo di questi tag non viene visualizzato dall'utente, ma è analizzabile dai crawlers e la sua importanza sta nel fatto che riesce a comunicare ai browser (o ad altri servizi web) informazioni specifiche riguardanti la pagina web in questione. Detto in parole più semplici, i meta tag "spiegano" i contenuti, la struttura e il layout della pagina in modo tale che il browser possa capirlo.

La struttura dei meta tag si presenta in questo modo:

```
<head><title> Tag Title </title> <meta name="description" content="Page  
Description Here"> <meta http-equiv="content-type"  
content="text/html;charset=UTF-8"> </head>
```

Inizialmente andrà inserito il titolo della pagina e in seguito la descrizione che consentirà ai motori di ricerca una più rapida comprensione dei contenuti di pagina, al fine di una migliore selezione.

Mentre ci sono ancora diversi dibattiti su quali siano i meta tag più utili e importanti per i motori di ricerca, è chiaro che essi, da soli, non costituiscano una soluzione magica per guadagnare un elevato *ranking* nei diversi motori di ricerca. Tuttavia, certamente i meta tag sono l'unico mezzo in grado di spiegare ai motori di ricerca i contenuti del sito, e quando vengono implementati non correttamente l'effetto negativo può essere sostanziale e definitivo. Assieme all'effettivo contenuto del sito Web, i titoli e le descrizioni delle pagine giocano un ruolo molto importante nella determinazione di come i motori di ricerca interpreteranno i siti internet. Una nota a parte ricorda che i titoli e le descrizioni verranno lette anche dagli utenti e quindi i tag devono essere comprensibili per tutti, non solo per gli strumenti meccanici. Conseguentemente, è molto importante utilizzare le parole chiave più pertinenti al sito e più in grado di stimolare la curiosità e l'interesse dell'utente.

Elementi di meta sono elementi HTML o XHTML usati per fornire informazioni su una pagina web utili per i motori di ricerca e gli utenti del sito web.

Tali elementi devono essere posizionati all'interno del tag `<head> </head>` di un documento HTML e sono considerati parte della sua intestazione⁸⁸. Questi elementi sono:

- tag title⁸⁹
- meta tag description

88 Maltraversi, 2012, pag. 124

89 Tecnicamente non si tratterebbe di un meta tag, ma viene comunemente associato ad essi dalla letteratura scientifica; nell'indicizzazione di un sito è ritenuta la componente più importante

- keywords⁹⁰

2.2.4 Title tag

Il title tag è un breve dichiarazione del contenuto di ogni pagina Web, simile a come all'interno di un libro il titolo di ogni capitolo ne spieghi la trama, e costituisce lo strumento più importante per quanto riguarda l'indicizzazione di un sito. Si è già parlato del funzionamento dei motori di ricerca: quando uno *user* effettua una *query*, in automatico il search engine scansionerà i title tag delle pagine Web, che sono il primo elemento visualizzato sia dai crawler che dagli utenti. Nel titolo di ogni pagina bisogna scegliere le combinazioni di keyword che meglio riescano a descrivere la pagina selezionata, onde ottenere un ottimo ranking.

Per quanto riguarda la lunghezza del titolo, diversi esperti raccomandano di mantenersi tra i 50 e gli 80 caratteri (spazi inclusi) per ottenere buoni risultati. La lunghezza accettata dai motori di ricerca può variare, Google ad esempio non visualizza più di 72 caratteri. Nel caso si scelga comunque di scrivere un titolo di lunghezza superiore allora è probabile che il sito sia inteso non per una fruizione umana quanto per i search engine, in quanto la scelta potrebbe apparire poco professionale e poco chiara, diminuendo le possibilità di visualizzazione da parte di potenziali clienti.

Un consiglio dato da numerosi esperti SEO, è di evitare l'utilizzo di titoli come "Home" o "pagina di benvenuto". Questi titoli, anche se concisi, non presentano chiaramente il contenuto della pagina e la loro genericità non consente al motore di ricerca, o all'utente, di capire subito l'argomento di cui si sta trattando. Poiché non viene espressa alcuna informazione specifica, un titolo di questo tipo non è in grado di fornire alcun valore SEO.

Oltre al fenomeno già spiegato di *keyword stuffing*, può succedere che Google autonomamente ritenga che le parole chiave scelte non siano adeguate e quindi provveda a sostituirle con altre da esso ritenute migliori. Non esiste alcuna garanzia volta ad impedire che questo possa verificarsi, l'unico modo per difendere il proprio title tag è assicurarsi che non sia generico o farcito di descrizioni superflue; deve essere chiaro e conciso. È importante avere titoli distinti per ogni pagina del sito, dato che ognuna di esse contiene informazioni diverse. Scegliere di mantenere i titoli tutti uguali porterebbe solo confusione all'utente, che sarebbe impossibilitato dal capire a prima vista le differenze tra le pagine. Questa operazione è importante anche perché facilita l'individuazione della pagina, da parte dell'utente, nell'eventualità che l'abbia inserita tra i preferiti. Anche i titoli stereotipati andrebbero evitati, si tratta di lunghi titoli che variano solo per un unico elemento di informazione, mentre per il resto si ripetono identici in ogni pagina.

⁹⁰ Il loro uso è oggetto di discussione continua da parte degli esperti, si spiegheranno nei paragrafi successivi la storia e gli sviluppi del keyword meta tag

Nel caso non venga riempito lo spazio title tag, la pagina verrà automaticamente segnalata come *untitled document* dai motori di ricerca, azzerando quello che è un importante fattore di ranking. Il title tag è quanto verrà visualizzato nei risultati di ricerca, se il titolo è molto chiaro e descrittivo la pagina Web avrà buone possibilità di posizionamento, ma attenzione a non utilizzare parole chiave estranee ai contenuti; nel momento in cui i *crawler* realizzano l'inganno si ricevono pesanti penalizzazioni. Sono sconsigliate anche le cosiddette *poison word*, si tratta di parole in conflitto con le utility del web (vd Bookmark, Link, Directory, Forum,..) e pur non causando una rimozione completa dall'indice, penalizzano comunque il posizionamento all'interno della SERP⁹¹. Normalmente è consigliabile inserire come prima parola la keyword più importante e poi, se vie è necessità di separare le parole, è meglio utilizzare la barra verticale, in inglese "pipe",|, piuttosto del trattino "-" che non induce Google ad archiviare i termini separatamente. Se le combinazioni di parole chiave scelte sono ripetute anche all'inizio del contenuto della pagina, allora esse sono doppiamente rafforzate nel momento della scansione.

In conclusione, il titolo dovrebbe essere abbastanza chiaro per gli spider ed accattivante per gli utenti in modo da garantirne il click, a scapito dei competitori diretti presenti nella SERP.

2.2.5. Meta description

Le descrizioni dei meta tag, assieme ai title tag, trasmettono l'argomento del sito. Si tratta di strumenti molto importanti per il raggiungimento del target di utenti desiderato. Per fare in modo che il motore di ricerca capisca esattamente di cosa parli il sito, non è sufficiente un buon titolo, bisogna aggiungere un'ottima descrizione. Anche in questo caso il posizionamento delle parole chiave non è da sottovalutare; se esse vengono posizionate all'inizio del tag, Google le marca come più rilevanti nelle ricerche. Tuttavia, questo non significa che si debba riempire la descrizione con elenchi di parole chiave, al contrario, dovrebbe leggersi come un frase. Una buona descrizione contribuisce ad aumentare la percentuale di clic sul sito web, in quanto non solo comunica agli utenti il contenuto della pagina, ma è la determinante che può spingerli a scegliere il proprio sito Web piuttosto che quello di un competitor con un ranking migliore. Come nel caso del tag title, anche in questo caso bisogna scegliere una descrizione con un testo "user-friendly" e conciso. I siti con descrizioni scarse o liste di keyword, otterranno meno clic e conseguentemente verranno fatti slittare in fondo ai risultati dai motori di ricerca a favore di altre pagine web più visualizzate.

Un webmaster può anche decidere di lasciare lo spazio dedicato alla descrizione della pagina vuoto, quando ciò si verifica i motori di ricerca riempiono in automatico lo spazio; in alcuni casi (come per

91 Maltraversi, 2012, pp. 123-128

esempio i blog) questa scelta è quasi consigliabile.

Una delle principali differenze tra tag title e meta description è la lunghezza. Generalmente viene suggerito di non eccedere le 12-15 parole, per un massimo di 140/160 caratteri⁹². Allo stesso modo del tag title, nel momento in cui un sito è formato da più pagine, per ognuna di esse andrebbe effettuata una descrizione differente e specifica. Descrizioni identiche per ogni pagina non sono utili per gli utenti; nel caso si abbiano pagine dinamiche è possibile generare automaticamente il tag prendendolo da parte del corpo della pagina. Se per qualche motivo si fosse impossibilitati dal creare descrizioni differenti per ogni singola pagina, bisognerebbe almeno cercare di definirne la priorità del contenuto; per esempio si possono creare descrizioni speciali per gli URL più critici, come la home page o le pagine più popolari.

Ad una buona descrizione meta, composta da frasi attraenti per gli utenti, potrebbe essere molto utile includere dati strutturali alla pagina. Ad esempio, se si sta pubblicizzando la pagina di una mostra da tenersi presso il Museo della Città, sarebbe senz'altro utile fornire informazioni collegate (date, costi, localizzazione, artisti esposti, ..). Tutte queste specifiche sottotitolo sono essenziali per i visitatori e se non si usassero dati strutturati è improbabile che Google le mostrerebbe nella descrizione.

Una meta description ben fatta potrebbe anche riunire dati sparsi tra varie pagine in modo da ottenere una sintesi coerente. Ad esempio, se si volesse scrivere una descrizione della mostra di Lorenzo Mattotti, dal titolo Favole e Misteri che si tiene presso la FAR Museo della Città di Rimini:

```
<meta name="Description" content="Nome Mostra: Favole e Misteri,
```

```
Artista: Lorenzo Mattotti, Categoria: Mostre, Prezzo: Ingresso Libero, Data Inizio: 24 maggio 2013,
```

```
Data Fine: 28 luglio 2013, Orari 16-23, Giorno Chiusura: Lunedì, Info: 0541.704416, Indirizzo
```

```
Mail: www.riminifar.it">
```

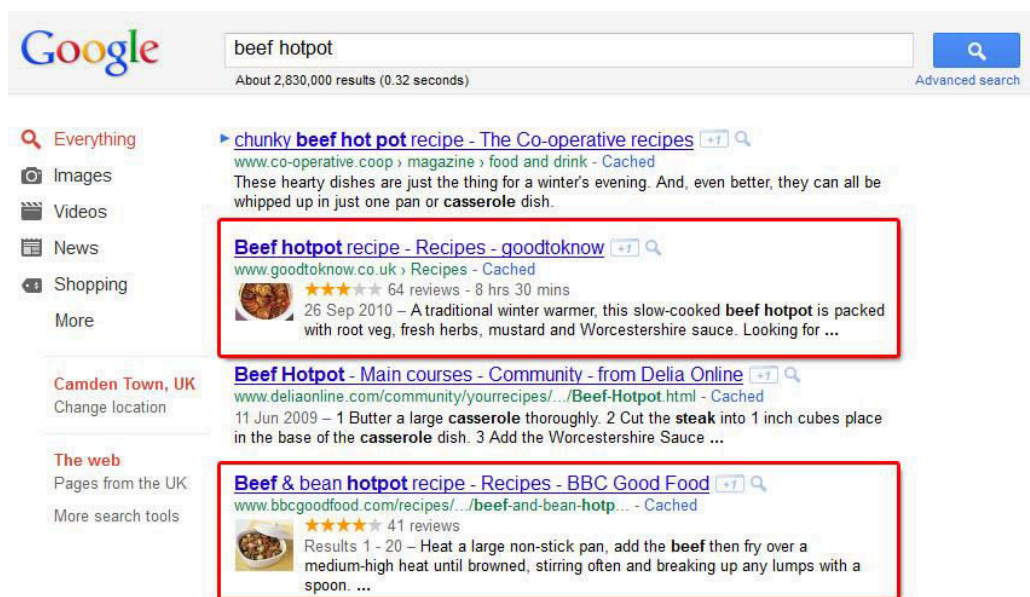
2.2.6. Google meta description: snippet e rich snippet

Nel caso specifico di Google, il motore di ricerca utilizza i cosiddetti “snippet” (traduzione italiana “frammento”) per descrivere una pagina web. Questi *snippet* vengono generati automaticamente quando un utente effettua una *query* usando una parola chiave e Google seleziona una serie di URL evidenziando non solo la parola in questione, ma anche quelle vicine onde comunicare al meglio il contenuto del sito Web. Nelle pagine strumento di supporto per Webmaster fornite da Google sono proprio i tecnici del search engine a spiegare come si possano ottenere *snippet* ricchi di

92 Maltraversi, 2012, p. 129

informazioni mediante l'utilizzo di dati strutturati. Citando quanto scritto: “Se Google capisce il contenuto delle pagine, è possibile creare *snippet* ricchi di informazioni e particolareggiati destinati ad aiutare gli utenti e incontrare le loro necessità. Ad esempio, lo *snippet* di un ristorante potrebbe mostrare alcune recensioni e la fascia di prezzo medio, mentre il frammento di una pagina di ricetta può mostrare il tempo totale di preparazione, una foto, e un *rating* della ricetta, mentre il frammento di un album musicale potrebbe elencare tutte le canzoni insieme con un link per poterle ascoltare subito. Questi “rich snippets” aiutano gli utenti a comprendere quando il sito è rilevante per la loro ricerca, e può comportare più clic alle tue pagine”⁹³ Per ottenere degli snippet molto ricchi di informazioni vengono dati tre suggerimenti che i webmaster possono facilmente implementare nel momento di costruzione del sito web. Il primo passo consiste nella scelta di un formato di markup, Google suggerisce l'utilizzo di microdati⁹⁴, ma anche microformati⁹⁵ e RDFa⁹⁶ sono accettabili. Viene anche sottolineato che non è necessaria una conoscenza approfondita di questi argomenti, ma è sufficiente una generale conoscenza del linguaggio di HTML.

Secondariamente il team di Google suggerisce di contrassegnare il contenuto della pagina. I “rich snippet” riconoscono e supportano diversi tipi di contenuti, come recensioni, persone, prodotti,



93 <https://support.google.com/webmasters/answer/99170>

94 I microdati costituiscono una modalità per etichettare il contenuto in modo da descrivere specifici tipi di informazioni, come eventi, dati di una persona, ecc. Ogni tipo di informazione descrive uno specifico elemento, come ad esempio una società, un evento o una recensione.

95 Citando la definizione data sulla pagina di supporto di Google, i microformati sono delle “convenzioni note come entità, utilizzati nelle pagine web per descrivere un tipo specifico di informazioni, come recensioni, eventi, aziende e persone”. Ogni entità ha le sue proprietà, ad esempio un'azienda avrà un indirizzo, una ragione sociale, un indirizzo e-mail, una partita iva, ecc.

96 Utilizzando la definizione della pagina di supporto ai webmasters di Google, RDFa sono un “modo per etichettare il contenuto di un sito in modo da poter descrivere uno specifico tipo di informazione, come la recensione di un ristorante, un evento o una persona. Questi tipi di informazioni sono chiamati entità o item. Esattamente come per i microformati ogni entità sarà dotata di un numero di proprietà”.

ricette, eventi, musica e inoltre Google riconosce i markup per contenuti video e li utilizza per migliorare i risultati di ricerca. Infine viene suggerita la verifica del codice di markup tramite lo strumento di prova dei dati strutturati; in questo modo Google può leggere ed estrarre i dati marcati. Questo procedimento è facilmente implementabile da un esperto webmaster, anche se è stato sottolineato dall'azienda che la marcatura di dati al fine di ottenere snippet più ricchi potrebbe non influenzare il posizionamento del sito tra i risultati di ricerca e non garantire l'utilizzo del markup.

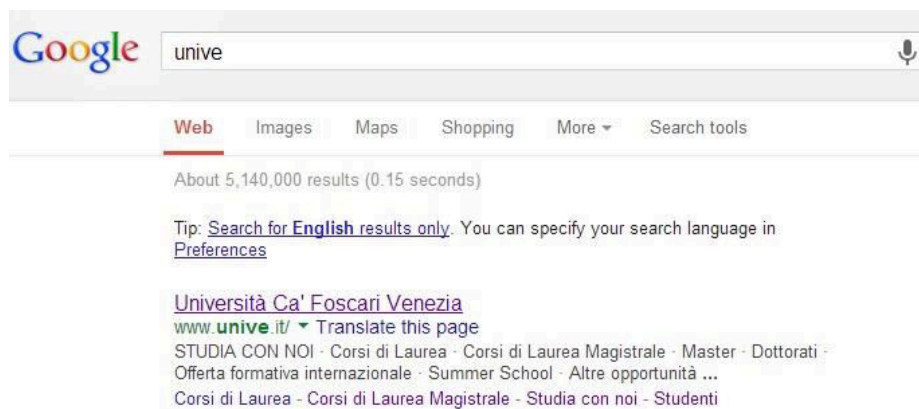
Riassumendo, su Google le descrizioni si formano in questo modo:

- Snippet & Descrizione ODP
- Snippet
- Meta Tag

Open Directory

Quando possibile, Google preferisce elencare le pagine utilizzando sia uno snippet che la descrizione della pagina, se essa è presente nella Open Directory⁹⁷.

Osservando i risultati di ricerca, la prima riga contiene il titolo della pagina mentre le successive costituiscono lo snippet scelto da Google in base alle parole che l'utente ha inserito durante la sua ricerca, visualizzate in grassetto. Successivamente, se la pagina web è inserita nelle ODP è possibile visualizzare anche il paragrafo dedicato alla descrizione.



Questa descrizione viene realizzata dagli editori di Open Directory, e la possibilità di controllare questo elemento è molto limitata per il proprietario della pagina web. Nel caso venga fornita una descrizione fuorviante o erronea è possibile contattare l'editore della categoria di Open Directory in questione e fare presente i propri motivi di contestazione. Se le motivazioni vengono ritenute

⁹⁷ L'Open Directory Project (conosciuto come ODP o anche come DMOZ, termine derivante dal nome del dominio originale directory.mozilla.org) è una directory di proprietà di AOL che contiene liste di siti organizzate secondo categorie di argomento. Questo progetto viene spesso usato da Google come database per i risultati di ricerca, e ciò in quanto i siti ivi catalogati sono analizzati e suddivisi non automaticamente, ma da schiere di volontari, garantendo affidabilità dei contenuti

effettive e reali allora vengono apportate le modifiche richieste. È possibile cliccare sul link della categoria e vedere se la pagina in questione risulta nella versione Google dell'Open Directory. Se una pagina non è presente all'interno delle Open Directory allora non appare nessun campo di descrizione. Se si vuole evitare che i search engine recuperino da DMOZ le descrizioni da posizionare nello snippet allora si possono usare dei meta tag opportuni come⁹⁸:

- Google: <meta name="googlebot" content="noodp">
- Bing: <meta name="msnbot" content="noodp">
- Yahoo! : <meta name="slurp" content="NOYDIR"/>

Può succedere che Google utilizzi un tag di meta descrizione per riempire lo spazio destinato allo snippet, nel caso non esista alcun testo consistente all'interno della pagina, ma ciò è molto raro.

Quando l'algoritmo di Google valuta come inaccurate le descrizioni le sostituisce in automatico con un'altra versione, anche se è chiaro che per una migliore utilità del sito Web dovrebbe essere il webmaster a spiegare ai potenziali utenti di cosa si occupa, piuttosto che lasciare il tutto a degli algoritmi meccanici.

2.2.7. Keywords Meta Tag: da usare o no?

Questo tag era utilizzato all'inizio della storia di Internet per il ranking dei siti Web; i primi motori di ricerca basati sull'uso di crawlers che utilizzavano meta keywords furono Infoseek, AltaVista e Inktomi (con il meccanismo di ricerca HotBot), nel 1996. Il successo di questo tag non durò a lungo per via dell'alta minaccia rilevata di spam,; ciò perché molti proprietari di siti Web inserivano parole fuorvianti riguardo le loro pagine o ripetevano continuamente keywords a scopo di ingannare i crawler sul contenuto del sito. Per questo motivo progressivamente i principali web crawlers smisero di analizzare il tag e i moderni motori di ricerca, come Google, non li hanno mai supportati. A sostegno di ciò si riportano le parole dello stesso Matt Cutts che dice: "Google does not use the keywords meta tag in web ranking"⁹⁹. Per questo motivo molti esperti¹⁰⁰ ritengono questo tag superato e praticamente inutile da implementare. È però fatto presente da altri¹⁰¹ che, nel dubbio, l'utilizzo di questo tag non è da eliminare in quanto altri search engines potrebbero tuttora farne uso. Uno dei motori di ricerca che ancora ne farebbe uso è Microsoft Bing, come aiuto contro il rilevamento di spam. Per evitare di danneggiare un nuovo website, la scelta migliore potrebbe essere quella di non inserire questo tag. Se invece si sta svolgendo un'analisi di un sito già attivo da tempo al cui interno sono presenti keywords meta tag non rilevate come spam, allora non c'è

98 Maltraversi, 2012, pag. 130

99 <http://tinyurl.com/tag-keyword>

100 Schachinger, 2012 ; Sullivan 2002 ; Goodman 2002

101 Maltraversi, 2012, pag. 131

motivo di rimuoverle tutte in una volta; si possono eliminare gradualmente, se non altro per ridurre il peso della pagina e i tempi di caricamento. Nel caso si scelga di inserire questo tag, allora è consigliabile l'utilizzo di massimo 40 parole chiave, evitando accuratamente ripetizioni, separandole con una virgola e uno spazio. Queste parole chiave, ai fine di una buona rilevanza, dovranno essere le stesse dei tag title, della description e del body della pagina.

2.2.8. Meta tag language

Questo breve tag è utilizzato per indicare la lingua della pagina web e può essere usato dagli strumenti di traduzione automatica, come Google Translate:

```
<meta http-equiv="content-language" content="it"102>
```

Se invece non si desidera che una pagina del sito non venga tradotta si dovrà utilizzare il meta tag:

```
<meta name="googlebot" content="notranslate">
```

Se non si vuole tradurre solo una porzione del testo allora il tag da utilizzare è:

```
<div class="notranslate"> testo non da tradurre </div>
```

2.2.9. Meta Tag Robots

Il compito di questo tag è quello di segnalare al crawler che sta scannerizzando un sito Web a quali pagine può accedere, quali link all'interno di esse può seguire e quali invece non deve indicizzare. È importante implementarlo correttamente perché se per errore si aggiunge un "noindex" tag, l'intero sito non verrà indicizzato. Se non viene inserito nessun meta robots allora i crawler indicizzeranno in automatico tutte le pagine del sito e seguiranno tutti i link presenti. Questo tag è utile nel caso in cui siano presenti contenuti duplicati, pagine protette o altre situazioni critiche per cui una determinata pagina non deve essere indicizzata dal motore di ricerca¹⁰³.

Esistono quattro implementazioni del Meta Tag Robots:

- 1) `<meta name="robots" content="index, follow">` richiesta di indicizzare la pagina nell'archivio del motore di ricerca e seguire tutti i link alle altre pagine del sito Web;
- 2) `<meta name="robots" content="index, nofollow">` richiesta di indicizzare la pagina, ma di non seguire tutti i collegamenti alle altre pagine;
- 3) `<meta name="robots" content="noindex, follow">` richiesta di non indicizzare la pagina, ma di seguire tutti i collegamenti alle altre pagine;
- 4) `<meta name="robots" content="noindex, nofollow">` richiesta di non indicizzare la pagina

¹⁰² Si sceglie "it" come content se la pagina è in italiano, altrimenti si usa "en" per l'inglese, ecc.; all'indirizzo <http://tinyurl.com/codici-iso> si può trovare l'elenco dei codici ISO utilizzabili

¹⁰³ Maltraversi 2012, p. 134

e di interrompere i collegamenti presenti sul sito tra le pagine.

L'ultimo tag, "noindex, nofollow" è utile quando un sito Web è sotto allestimento, allora lo sviluppatore vuole evitare che gli utenti si colleghino a una pagina non ancora terminata. È però un errore piuttosto comune quello di dimenticarsi poi di rimuovere il tag, una volta terminata la costruzione del sito.

2.2.10. Heading Tag

I tag di intestazione si usano per organizzare i testi della pagina, dando più o meno rilevanza alle frasi racchiuse al loro interno. Ci sono sei tipologie utilizzabili, dal tag <H1> che è quello indicante l'argomento della pagina, fino al tag <H6>, che è quello meno significativo. Tramite questo strumento è possibile segnalare al crawler, agli utenti e ai browser quali siano le frasi più importanti da visualizzare nella pagina Web. Sostanzialmente il loro compito è quello di aiutare a dare una priorità ai contenuti e a facilitare l'indicizzazione della pagina.

L'utilizzo dei tag è regolato da alcune direttive:

- gli heading tag devono essere utilizzati in maniera decrescente, da <H1> a <H6>;
- è possibile usare una sola intestazione H1 per pagina;
- se si intendono utilizzare tutti e sei i livelli bisogna procedere coerentemente, senza effettuare alcun salto;

I sei livelli possono essere spiegati in questo modo:

- 1) Tag <H1> per il titolo della pagina e indica il topic
- 2) Tag <H2> e <H3> per i sottotitoli della pagina
- 3) Tag <H4> e <H5> per i testi della pagina
- 4) Tag <H6> per note, approfondimenti e diritti d'autore

Dato che i crawler effettuano la scansione dei siti Web procedendo da destra verso sinistra e dall'alto verso il basso, sarebbe utile inserire il tag H1 in alto a destra.

2.2.11. Ottimizzazione della pagina

I siti Flash

Per quanto riguarda l'ottimizzazione della pagina, anzitutto si voglio segnalare due suggerimenti di cui si è letto svariate volte durante la ricerca sulle tecniche SEO da implementare sul proprio sito Web. Quello che viene sconsigliato dalla maggioranza degli articoli specializzati nell'argomento¹⁰⁴ è di utilizzare due elementi. Il primo è costituito dalle cosiddette "splash pages" (spesso consistenti di un'animazione Flash) e il secondo consiste genericamente di tutti i siti

104 Search Engine Journal

Flash. Questo perché Flash, pur essendo esteticamente piacevole, spesso con artwork interattivi, ha il grosso difetto di non essere assolutamente rilevante ai fini del ranking. Al contrario, nonostante pare che ci possano essere futuri possibili miglioramenti, al momento i motori di ricerca non sono in grado di leggere Flash. Si possono scrivere dei contenuti di pagina molto buoni e dall'elevato valore SEO, ma, se inclusi in un sito Flash, allora per il motore di ricerca è come se il sito fosse invisibile, tagliando fuori una consistente porzione di possibili fruitori. Allo stesso modo le pagine splash (una volta molto diffuse come una sorta di “anticamera” del sito, con animazioni flash) sono attualmente sconsigliate. Molti utenti potrebbero non gradire il dover fare quell'extra clic per poter accedere ai contenuti principali del sito o, peggio ancora, se non si fornisce uno “skip button” (ovvero un pulsante per saltare l'introduzione) si costringono i visitatori a visualizzare l'intera animazione, rischiando che una buona percentuale di essi abbandoni il sito prima ancora di visualizzarne i contenuti. Oltre a tutti questi disagi, se non si fornisce un link html conducente al sito, anche i motori di ricerca non potranno proseguire nella scansione della pagina (a meno che non si sia preparato un file sitemap.xml). L'utilizzo di Frames, Flash e AJAX comporta quindi il problema di non riuscire ad effettuare collegamenti tra le pagine. Per questo motivo, se necessariamente si vuole avere la pagina principale con animazioni Flash, si consiglia di posizionare il testo e i link di navigazione sotto alla cartella. Ancora meglio si potrebbe limitare l'animazione flash solo all'intestazione, perché si tratterebbe solo di un'aggiunta alla pagina Web, non di un elemento interamente separato.

Le breadcrumbs

Secondariamente è importante che ogni pagina sia ottimizzata sia a livello di title che di description e body. A questo proposito, sarebbe utile fornire all'utente delle “briciole di pane” onde evitare di perdersi all'interno del sito. Queste briciole sono le cosiddette “breadcrumbs” e consistono in delle indicazioni in grado di suggerire all'utente la sua posizione rispetto alla struttura del sito. Se per esempio si stanno visualizzando le offerte formative dell'Università Ca'Foscari, allora la pagina web in alto a sinistra fornisce una tracce, per mostrare all'utente dove si trova rispetto alla home page. In questo caso si visualizzerà:

home > studia con noi > corsi di laurea

dove le parole chiave “home” e “studia con noi” hanno due link di rimando alle rispettive pagine¹⁰⁵.

105 Taverniti 2011, pp. 50-51

2.2.12. Ottimizzazione delle immagini

Ridimensionamento e formati

Il modo in cui gli utenti visualizzano una pagina Web e il modo in cui la vede un crawler sono due cose molto diverse. Partendo da questo presupposto ci sono numerose considerazioni da fare. Anzitutto se si sceglie di inserire delle immagini all'interno del proprio sito (cosa molto consigliabile sia per i fattori di ranking, sia per l'attrattività che le immagini esercitano sugli utenti) bisogna avere cura di prendere alcune accortezze. Tra il testo inserito in una pagina e le immagini ci deve essere un equilibrio, non solo estetico (fattore in questo momento secondario) ma di peso. Le immagini hanno un peso nettamente superiore rispetto alle porzioni di codice e quindi una loro diffusione non controllata può portare numerose conseguenze negative. È bene ricordare infatti che molti crawler, mentre scansano una pagina, fermano la loro attività non appena raggiungono una determinata quantità di byte, in molti casi attorno ai 100 Kbyte. Raggiunto questo valore il crawler non andrà oltre nella scansione di quella determinata pagina, escludendo potenzialmente testi importati. Oltre a ciò il peso delle immagini influisce sui tempi di caricamento. Molti utenti non sono disposti ad aspettare oltre gli 8-10 secondi per visualizzare una pagina; dopo questo breve lasso di tempo è probabile che scelgano di chiuderla e passare ad un altro URL. Già dal novembre 2009, Google aveva annunciato¹⁰⁶ che la velocità di caricamento sarebbe diventato un fattore importante di ranking e ciò è stato confermato dagli studi più recenti. Esistono molte tecniche volte alla riduzione dell'impatto delle immagini digitali, partendo dalla semplice compressione. Secondariamente è bene considerare che il formato Jpeg non è sempre necessario per ogni immagine presente su una pagina. Spesso la scelta del formato GIF può costituire un'alternativa migliore; questo formato difatti è nettamente preferibile sui piccoli elementi grafici come icone o pulsanti. La mappa dei colori su cui si fonda il GIF è più ristretta rispetto a quella del Jpeg e, in caso di compressione, le immagini non si rovinano. Per questo motivo se la resa estetica delle immagini da caricare non è fondamentale, sarebbe meglio selezionare il formato GIF¹⁰⁷.

Alt Tag

Un altro aspetto importante delle immagini, oltre al controllo del loro peso, è la loro indicizzazione e ricercabilità su un search engine. Non bisogna dimenticare che Google Image Search è il motore verticale di Google per la ricerca delle immagini sul Web, ed è uno strumento fondamentale per aumentare la visibilità della pagina da cui le singole immagini vengono tratte.

106 <https://www.pdxtc.com/wpblog/google/google-shows-us-how-to-go-faster/>

107 Vasta, 2009, pp. 251-253

Per questo motivo è bene iniziare facendo in modo che l'immagine abbia un URL in grado di descriverne il contenuto. Secondariamente è consigliato inserire l'attributo Alt, elemento del tag con una serie di informazioni specifiche e accurate descrizioni delle immagini presenti sulla pagina Web. Così facendo il crawler sarà in grado di attribuire un vero e proprio "senso" all'immagine, mostrandola a chi sta effettuando una ricerca correlata anche solo con una porzione delle parole chiave. Un altro effetto importante è che se anche l'immagine non fosse più disponibile, comunque il testo alternativo¹⁰⁸ verrebbe trascritto. Quando l'attributo alt viene inserito su un'immagine contenente un collegamento ad un altro sito, il testo alternativo diventa anche l'anchor text del link e può essere quindi sfruttato per aumentare la link popularity della pagina¹⁰⁹.

Lo strumento Lynx

Si è detto di come i motori di ricerca non siano in grado di vedere immagini o altri elementi grafici, sebbene siano in grado di riconoscerne la presenza sulle pagine. Per aiutare il webmaster a capire se la struttura della pagina Web sia adeguata o meno esiste uno strumento chiamato Lynx. Si tratta di un browser "cieco", ovvero funzionante esattamente come un crawler, che analizza solo il testo di una pagina. Come spiegato da Davide Vasta: "Lynx¹¹⁰ permette di esaminare la struttura delle pagine, basandosi unicamente sui testi, sui link e sui tag Alt delle immagini"¹¹¹. Potrebbe essere quindi molto utile aiutarsi con questo programma per verificare che i testi non appaiono confusionari o ripetitivi, col rischio di compromettere il ranking del sito.

2.2.13. Scrittura delle URL

Il migliore consiglio che si può dare per quanto concerne la scrittura delle URL è quello di renderle il più "search-engine-friendly"¹¹² possibile, utilizzando come testo delle parole chiave chiare e correlate al proprio sito. Evitare gli URL dinamici, anche se contengono parole facilmente identificabili. Una URL dinamica è una pagina che all'indirizzo Web contiene caratteri speciali, definiti variabili. Normalmente i siti che fanno uso di queste URL si appoggiano a dei database contenti tutte le info e a seconda di quanto richiesto dagli utenti mostrano risultati diversi. Una volta i motori di ricerca faticavano a scansionare e indicizzare gli URL dinamici perché le combinazioni apparivano numerosissime; recentemente però Google ha dichiarato di aver risolto il problema. Altri search engine potrebbero però avere ancora difficoltà, per questo motivo si

108 Alt tag deriva dall'abbreviazione del termine "alternativo"

109 Vasta 2009, pp. 262-263

110 Il prodotto è scaricabile all'indirizzo <http://lynx.isc.org/>

111 Vasta, 2009, pag. 251

112 <http://www.businessinsider.com/10-basic-seo-tips-everyone-should-know-2010-1?op=1>

consiglia comunque l'utilizzo di URL statiche. Questo perché una URL statica può contenere parole chiave di cui si può beneficiare e inoltre è di più semplice da ricordare per gli utenti nel caso si scriva l'indirizzo su un supporto cartaceo e l'utente debba inserirlo nella barra di ricerca a memoria¹¹³.

2.2.14. Creazione della Sitemap

Nel primo capitolo, all'interno del paragrafo dedicato al processo di scansione, si è detto di come Googlebot sia il meccanismo tramite cui Google riesca ad accedere alle pagine Web. Un altro metodo per favorire l'indicizzazione di una pagina Web è la creazione di una Sitemap. Viene consigliata in letteratura¹¹⁴ la creazione di due Sitemap, una per gli utenti e l'altra, in formato XML, per i motori di ricerca. La prima sarà una pagina raffigurante la struttura del sito, chiaramente sempre da tenere aggiornata nel tempo, che funge proprio da mappa in modo che i visitatori non possano "smarrirsi" durante la loro perlustrazione del sito.

La Sitemap come file XML contiene tutti gli URL di un sito, tutte le pagine, i post, ecc. Sostanzialmente costituisce un valido supporto per illustrare ai crawler tutti i contenuti del nostro sito, i collegamenti, le pagine. Include anche la date delle ultime modifiche, comunicando la frequenza di aggiornamento ed è anche possibile inserire un numero (che va da 0 a 1) riguardante la priorità delle pagine a parere del Webmaster più rilevanti. Il suo utilizzo facilita la scansione e indicizzazione di un sito Web; in poche parole, meno clic sono necessari per arrivare ad una determinata pagina su un sito Web, meglio è. Al sito ufficiale del protocollo Sitemaps¹¹⁵ si possono trovare tutti gli standard di implementazione, come il fatto che ogni map deve essere contenuta tra i tag `<urlset>` `</urlset>`, inserire le voci `<url>` per ogni URL come tag XML principale, ecc.¹¹⁶ In conclusione, grazie a questo sistema, è possibile segnalare in un unico momento tutte le pagine che compongono un sito Web, sia ai search engine, che al visitatore¹¹⁷.

2.3. Analisi Esterna (off-page optimization)

Tra le tecniche SEO riguardanti l'ottimizzazione off-page troviamo quelle che consentono al sito un mantenimento del ranking duraturo nel tempo¹¹⁸.

2.3.1. Link popularity

113 Taverniti, 2011, pp.53-54

114 Maltraversi, 2012, pp. 191-193

115 <http://www.sitemaps.org/it/>

116 Maltraversi, 2012, pp. 191

117 Taverniti, 2011, pag. 26

118 Maltraversi, 2012, pag. 74

È un fattore di straordinaria importanza, alla base del funzionamento del PageRank di Google, ed essenziale per una buona diffusione e popolarità della pagina Web. Nessuna delle tecniche finora elencate nella on-page optimization può avere successo se non è accompagnata da un buon sviluppo di link popularity. Questo fattore è molto difficile da conseguire, ma sono sempre di più le tecniche attualmente utilizzabili in grado di pubblicizzare il website. Spiegando brevemente il meccanismo, più siti Web hanno link di collegamento ad una determinata pagina, più i motori di ricerca la considerano affidabile e rilevante ai fini di medesime ricerche svolte da altri utenti. Sostanzialmente, il posizionamento dei siti Web nella SERP di Google, per determinate keywords, avviene proprio grazie alla link popularity. L'obiettivo dei search engine è quello di interpretare e fornire agli utenti quelle pagine a cui potranno essere maggiormente interessati. Dato l'enorme numero di pagine presenti online la scelta può essere significativamente aiutata dai link consigliati dai più popolari website dall'argomento correlato a quello della ricerca. I motori di ricerca operano continui riscontri per valutare la coerenza dei contenuti degli URL consigliati; i programmatori sono consapevoli del fatto che gli abili webmaster possono aggirare il meccanismo forzando la popolarità dei siti. Per questo motivo vengono svolte delle analisi della link popularity (link analysis) in grado di valutare il numero totale di inbound link qualitativi; cioè a differenza della link popularity, data dal numero totale dei siti con link di collegamento a una determinata pagina, qui si valuta la qualità stessa dei link. Sempre più motori di ricerca fanno uso di algoritmi in grado di discernere quei siti che offrono hosting gratuito a numerose pagine Web, per poi linkarle a se stessi. Per questo motivo recenti studi hanno mostrato come search engine non valorizzino le pagine Web ospitare in aree gratuite. Oltre a ciò si ricorda che gli anchor text dei link sono molto importanti; bisogna avere dei link di collegamento contenenti parole chiave rilevanti e contestuali a quelle del proprio sito. Pochi link derivanti da siti Web affidabili e trattanti lo stesso business del sito che si vuole collegare contano molto di più dell'averne tanti link da siti dagli argomenti più disparati. È per questo essenziale non forzare la link popularity e nel caso si venga contattati da altri siti è bene accertarsi della loro validità prima di accettare o offrire collegamenti¹¹⁹.

2.3.2. Link Building

Etimologicamente il link building fa riferimento, letteralmente, alla “costruzione di link”. Questo perché, onde poter dirigere più traffico al proprio sito Web, è necessario sviluppare relazioni con altri siti. Esistono diverse tecniche attualmente sfruttate dagli esperti SEO, aventi come obiettivo quello di aumentare la visibilità e il bacino di utenza. Essendo questo campo in fase di grande

119 Vasta, 2009, pp. 264-265 ; http://www.googlerank.it/Guide_SEO/Fattori_Ranking_Esterni/Link-Popularity.asp

sviluppo, pare rilevante menzionare che molte strategie sono di derivazione di marketing tradizionale e di *link baiting* o marketing virale, ovvero l'investire nei contenuti in modo da generare tanti link spontanei, dettati dalla curiosità e interesse degli utenti.

Scambio, integrazione e vendita di link

La prima tecnica attuata in questo campo è sicuramente quella dello scambio naturale di link, avente come conseguenza una maggiore visibilità per tutti i siti coinvolti. Il funzionamento è apparentemente semplice: basta scegliere un sito (della cui validità e rilevanza si è certi) e mettersi in contatto con il webmaster, chiedendo l'inserimento all'interno di una loro pagina di un link diretto ad una propria pagina. Allo stesso modo, sul proprio sito, si può procedere all'inserimento di un link all'altra pagina e così entrambi i siti gioveranno di questa relazione. Nel momento in cui avviene il contatto, è bene specificare al webmaster dell'altro sito le keywords di preferenza da inserire come anchor text, in modo che siano simili alle parole chiave contenute nei titoli e descrizioni della pagina del proprio sito. È consigliabile di evitare contenuti come: "visitare il sito" o "link here" per via dei testi privi di valori semantici e ininfluenti sulla link popularity. Bisogna anche evitare di collegarsi alle già citate link farms, i cui comportamenti negativi potrebbero influire negativamente sul proprio ranking. Oltre alle link farms i motori di ricerca penalizzano un sito quando esso fa parte di un "circuito chiuso nel web" attraverso i link¹²⁰. Questo significa che se alcuni webmasters si accordassero per scambiarsi i link tra i loro siti e basta, non otterrebbero risultati positivi perché gli algoritmi del motore di ricerca glielo impedirebbero. Oltre ai link possono effettuarsi degli scambi anche di banner, per questo motivo sarebbe sempre utile tenerli pronti ed essere disponibili ad integrare nel proprio sito quelli di altre pagine, dei nostri medesimi contenuti e della cui validità, appunto, si possa essere certi.

Secondariamente allo scambio di link, una delle tecniche base utilizzate dagli esperti SEO è quella dell'integrazione di link interni al proprio sito; in questo modo si può aumentare il traffico dati alle singole pagine. Dovrebbe essere procedura standard quella di collegarsi sempre al proprio archivio quando si creano dei nuovi contenuti. L'anchor text, come sempre, deve essere "search-engine-friendly" ovvero contenente parole chiave molto rilevanti e pertinenti alla pagina, in modo da aumentarne la probabilità di comparsa quando altri utenti effettuano ricerche con le stesse keywords¹²¹. Dopo scambio e integrazione sicuramente ai più sarà nota la possibilità di effettuare l'acquisto di link. L'obiettivo iniziale che ha portato allo sviluppo in Italia di questa tecnica era quello di aumentare il PageRank delle pagine, senza preoccuparsi di quanto effettivamente popolari

120 Taverniti, 2011, pag. 79

121 <http://www.marketingvox.com/how-to-9-basic-seo-tips-037438>

esse potessero diventare. Il mercato dei link era nel pieno della sua attività, si può immaginare il giro di denaro dietro alla vendita di link, fino a che Google, in seguito a svariate segnalazioni, è giunto alla decisione di penalizzare chi facesse uso smoderato di questa tecnica. La penalizzazione non ha avuto conseguenze a livello di posizionamento, difatti è stato colpito solo il punteggio di PageRank di quelle pagine che fornivano link in uscita, senza preoccuparsi della mancata correlazione di argomenti e senza il rel="no follow". Ad ogni modo se si sceglie di comprare link, è sconsigliabile di fare delle "prove" limitate nel tempo. Il valore dei link aumenta col loro tempo di permanenza su una pagina, comprarli per un mese per poi farli scadere non porta buoni risultati. Bisogna usare anchor diverse per ogni link, non accettare dei prezzi esorbitanti solo perché il sito che li propone ha un buon PageRank, non è affidabile. Ancora meglio, non comprare link da siti non tematicamente correlati al nostro¹²². Un sito a cui fare riferimento nel caso si decida di procedere con l'acquisto di un pacchetto di link è www.alverde.net¹²³.

Inserimento in Directory

Si è già spiegato nei paragrafi precedenti in cosa consistano esattamente le Directory, a questo punto è utile riprendere il discorso contestualmente al potenziamento di popolarità di una pagina Web. Difatti internet è pieno di web agency più o meno note e più o meno affidabili che propongono di inserire il proprio sito all'interno di svariate directory. L'inserimento normalmente dovrebbe essere gratuito, ma esistono delle situazioni per cui potrebbe risultare a pagamento. Questa tecnica, pur avendo avuto un peso importante in passato, attualmente ha un ruolo decisamente meno incisivo, dovuto al fatto che le principali directory (DMOZ, Yahoo!, Zeal, ..) sono rimaste identiche a come erano nel loro periodo di massima gloria, e quindi hanno perso gradualmente rilevanza. Se comunque si decide di effettuare un inserimento è meglio verificare prima la qualità dei link in uscita, onde evitare penalizzazioni dai search engine. Grande attenzione va posta nella correlazione tra anchor text dei link in uscita e il titolo della pagina, ci devono essere le medesime pagine chiave¹²⁴.

Inbound Marketing

Obiettivo dell'inbound marketing è quello di far sì che, a differenza del marketing tradizionale, sia proprio l'utente ad essere interessato all'azienda e a mettersi in contatto con essa¹²⁵. Nell'outbound marketing, quello tradizionale, si utilizzano una svariata serie di metodi molto più invadente, per cui

122 Taverniti, 2011, pp. 86-87

123 Maltraversi, 2012, pag. 319

124 Taverniti, 2011, pp.80-83

125 Maltraversi, 2012, pag. 75

si cerca di convincere i potenziali clienti, tramite annunci televisivi, o pubblicità, a scegliere e a interessarsi a quello offerto dalla propria organizzazione. Qui invece si cerca di creare dei contenuti (di svariata forma e supporto) che i clienti desiderino vedere. Si vogliono costruire delle relazioni tra aziende e clienti, tramite cui ci si vuole far conoscere e creare un rapporto più personale, che si sviluppi in una sorta di fedeltà derivante dalla familiarità ottenuta restando presenti quotidianamente nella vita dell'utente.

Article marketing e comunicati stampa

Queste tecniche si adattano molto bene nel caso il sito che si intende analizzare abbia argomenti accademici. La tecnica è molto datata e fa riferimento alla pubblicazione di un buon articolo che poi, in seguito alla condivisione, porta fama e notorietà per un determinato argomento. Attualmente alcuni considerano la tecnica come una pubblicazione ripetuta dello stesso articolo su piattaforme diverse, al solo scopo di ottenere backlink da diversi siti. Se l'articolo scritto è davvero di buon contenuto, la condivisione da parte degli utenti dovrebbe avvenire spontaneamente, su tutti i mezzi di diffusione web. Può essere d'aiuto tenere un blog aggiornato e segnalare i propri articoli solo ai siti specializzati che probabilmente vorranno pubblicare solo inediti, costringendo o a modificare il pezzo o a condividerlo su di un solo canale. Se non si è in possesso di tali mezzi (anche se tramite l'utilizzo di strumenti completamente gratuiti come YouTube, Google Universal Search ecc. si possono ottenere comunque ottimi risultati) conviene contattare chi possiede forum e blog molto in vista e chiedere loro di pubblicare l'articolo specializzato¹²⁶. Per quanto riguarda i comunicati stampa, essi vanno scritti professionalmente, rivolti ad un utente interessato, presumibilmente capace dell'argomento. Il comunicato stampa¹²⁷ è soggetto a regole molto severe, come un determinato numero di caratteri, forma e contenuti appropriati e inediti e sono abitualmente soggetti a controllo da parte di esperti prima di venire pubblicati¹²⁸.

Utilizzo del SMO

L'utilizzo di Social Network rientra in una moderna branca di studio denominata SMO, ovvero Social Media Optimization. I social media costituiscono un nuovo mezzo di interazione, oramai divenuto necessario, per rimanere in contatto coi clienti quotidianamente, trovarne di nuovi e diversificare le proprie fonti di traffico; questa tecnica, come visto nello studio sui Google Ranking

126 Una lista di siti di Article Marketing viene fornita dall'autore Giorgio Taverniti a questo link:

<http://liste.giorgiotave.it/article-marketing>

127 Una lista di siti di comunicati stampa sempre fornita dall'autore Giorgio Taverniti a questo link:

<http://liste.giorgiotave.it/comunicati-stampa>

128 Taverniti, 2011, pp. 84-86

Factors, ha un peso rilevante ai fini SEO. Il motto fondamentale è: “Essendo presenti in più luoghi, avremo più probabilità di attirare nuovi clienti”¹²⁹. La presenza diffusa di un determinato marchio su vari supporti va coordinata al target e al luogo all’interno del quale si opera. L’obiettivo ultimo di questa ottimizzazione è quello di incoraggiare i vari utenti a condividere i contenuti del proprio sito, in modo che diventino essi stessi mezzo di diffusione. Per fare in modo che ciò possa accadere bisogna produrre contenuti diversificati in base alle varie fasce coinvolte dai diversi mezzi, riuscire a far parlare di sì incuriosendo e creando contenuti di valore. Oltre a ciò, è basilare l’aggiornamento continuo dei social, l’essere costantemente presenti nelle vite degli utenti, spingendoli al dialogo per capire quali potrebbero essere i loro interessi e per determinare se le direzioni di sviluppo scelte siano quelle corrette. In letteratura ci si riferisce allo sviluppo del SMO come allo “sviluppo delle 4C”¹³⁰: contenuto, contesto, connettività e community. Il contenuto deve essere filtrato, aggregato e reso disponibile all’interno di un contesto che lo rende utile per gli utenti che desiderano mettersi in contatto con altre persone, prodotti e brand attorno della quale riescano a sviluppare una comunità distintiva. Quando si parla di Social Media si va dai social network come Twitter, Facebook e LinkedIn ai Social Bookmark come StumbleUpon, Delicious e Digg. Lo sviluppo di questi network è incoraggiato ai fini della creazione di relazioni e dell’aumento di contenuto di qualità del proprio sito.

Banner advertising e e-mail marketing

Il banner advertising consiste in una sorta di campagna pubblicitaria tabellare, per cui si individuano dei siti target su cui sviluppare la propria campagna. A differenza delle campagne di keyword advertising, acquistabili tramite cost per click, le campagne banner si acquistano in modalità cpm (costo per mille impressions); bisogna minimizzare i costi di acquisizione e successivamente monitorare il comportamento degli utenti in seguito alla campagna oltre a valutarne benefici a livello di branding.

A proposito di e-mail marketing invece, è necessario dire che esistono varie tipologie di e-mail, tra queste si ricordano quelle transazionali, inviate in automatico dopo che l’utente ha svolto una determinata azione online, come un acquisto o una registrazione. C’è poi il direct e-mail marketing (DEM) che consiste nell’invio di pubblicità di determinati prodotti e servizi. Esistono infine le newsletter, che sono messaggi inviati periodicamente in automatico agli iscritti per comunicare, come suggerisce il nome stesso, eventuali novità o cambiamenti all’interno di un sito. Gli elementi base in qualsiasi azione marketing svolta tramite e-mail sono chiaramente l’acquisizione delle liste

129 Maltraversi, 2012, pp. 382-383

130 Maltraversi, 2012, pag. 383

di indirizzi e-mail, per i quali ci vuole tempo e conoscenza degli utenti (senza dimenticare le norme vigenti sulla privacy¹³¹), seguito dall'invio di messaggi che siano effettivamente rivolti all'utente che li riceverà, si pensi alle campagne marketing di Amazon che, memore degli acquisti svolti da un utente, focalizzerà il suo DEM sugli articoli di maggiore interesse¹³².

Creazione del brand

Si è visto nel paragrafo dedicato ai Google Ranking Factors, come l'utilizzo del fattore brand possa favorire una determinata pagina Web rendendola esente da penalizzazioni che altri siti rischierebbero invece di subire. Questo obiettivo è raggiungibile tramite azioni di partnership e co-marketing, tramite accordi con portali contenutistici specializzati e web-zine. Tramite l'inserimento di un proprio servizio all'interno di più sezioni di un sito partner ben noto e già seguito da molti utenti si possono ottenere ottimi risultati. La creazione del brand è necessaria per avere successo in rete e un buon modo per svilupparla è chiedersi proprio cosa possa offrire il proprio sito di unico e interessante per gli utenti. Bisogna entrare su social media e servizi simili con un chiaro piano di marketing, onde evitare fatali feedback loop, passaparola in negativo che possono stroncare decisamente l'organizzazione in pochissimo tempo. A questo proposito, siti che permettono di monitorare la reputazione sui vari mezzi social, sono www.peakyou.com e www.pipl.com, ma anche www.onlineidcalculator.com che consente una verifica più generale della propria della propria fama e considerazione online. Successivamente, tramite una buona campagna di strategic inbound marketing, è possibile ottenere una forte identità di marchio, la fedeltà da parte dei clienti, che desiderano costantemente essere informati e aggiornati su quanto si sta facendo, con un conseguente ottenimento di link building e popularity; il tutto contribuisce a rafforzare il proprio sito agli occhi di algoritmi che ne valutano la qualità, quali Google Penguin.

131 Norme rintracciabili a questo indirizzo: <http://www.garanteprivacy.it/garante/doc.jsp?ID=432448>

132 Maltraversi, 2012, pp. 337-340

IL SITO DEI MUSEI DELLA CITTÀ DI RIMINI

Gli obiettivi di questa analisi sono quelli di migliorare la visibilità e potenzialità del sito dei Musei della Città di Rimini, partendo dal presupposto che, al giorno d'oggi, l'implementazione di tecniche SEO sia essenziale per raggiungere questi risultati. È opinione di chi scrive che l'attuale valore SEO del sito è estremamente limitato, per poter raggiungere un buon posizionamento nei risultati di ricerca è necessario effettuare delle migliorie in diversi aspetti del website. Nell'effettuare l'analisi del sito si è scelto di ricorrere all'uso di alcuni programmi di SEO Analytics, quali WooRank.com, SeoSiteCheckup.com., SearchCommander.com. e Alexa.com. Questi siti contengono dei software utili per effettuare delle website review che sono preziose ai fini di migliorare il marketing online di business di piccola portata. È possibile effettuare dei test gratuiti immettendo l'URL del sito di cui si vuole effettuare l'analisi, mentre per poter accedere a risultati approfonditi e a azioni bisogna iscriversi e, dopo una settimana di prova gratis, si procede al pagamento del primo mese di analisi, che normalmente si aggira attorno al centinaio di euro. Questi sono i metodi con cui si è effettuata l'analisi e la loro utilità è stata essenziale nell'ottenimento di report diagnostici e rappresentanti nella maniera più precisa e veritiera, la situazione del sito dei musei riminesi. È possibile fare già una piccola anticipazione sui risultati finali d'analisi, infatti il giudizio più diffuso su questi software di SEO analytics è che il sito museale risulta sufficientemente ottimizzato, ma c'è ancora tanto spazio per poterne migliorarne la visibilità, miglioramento che si tradurrebbe in termini di maggiore redditività. In fondo al capitolo si riporta una figura riassuntiva dell'intera analisi con il punteggio assegnato al website da WooRank.com, software in questo momento leader nel settore SEO Analytics.

Si intende adesso procedere all'analisi dei risultati ottenuti dai report diagnostici, per quanto riguarda la loro presentazione si è pensato di riprendere la struttura proposta all'interno del secondo capitolo, ovvero dividendo l'attività di studio in analisi di mercato, analisi dei fattori on-page quindi interna al sito e analisi di fattori off-page, esterni alla struttura del sito e riguardanti la visibilità sul web.

1. Analisi di mercato

Lo strumento Alexa.com ha suggerito durante la sua analisi la presenza di alcuni siti con nomi simili a quelli del museo (fornendone anche il rispettivo global rank), esattamente come avrebbe fatto nell'analisi di un sito di una compagnia specializzata in un determinato business.

Categories with Related Sites	
This site is not categorized	
Sites with similar names	Global Rank
1. museindia.com	1,233,111
2. musei.it	2,507,723
3. museicapitolini.org	1,022,860
4. museiciviviceneziani.it	29,718,558
5. museiincomuneroma.it	807,381
More	

Appare più che evidente che però, nel caso del sito di un museo, l'analisi di mercato non appare essere svolta esattamente come in una situazione di profitto, per cui un'azienda può identificare i suoi diretti competitors e sviluppare una strategia di conseguenza.

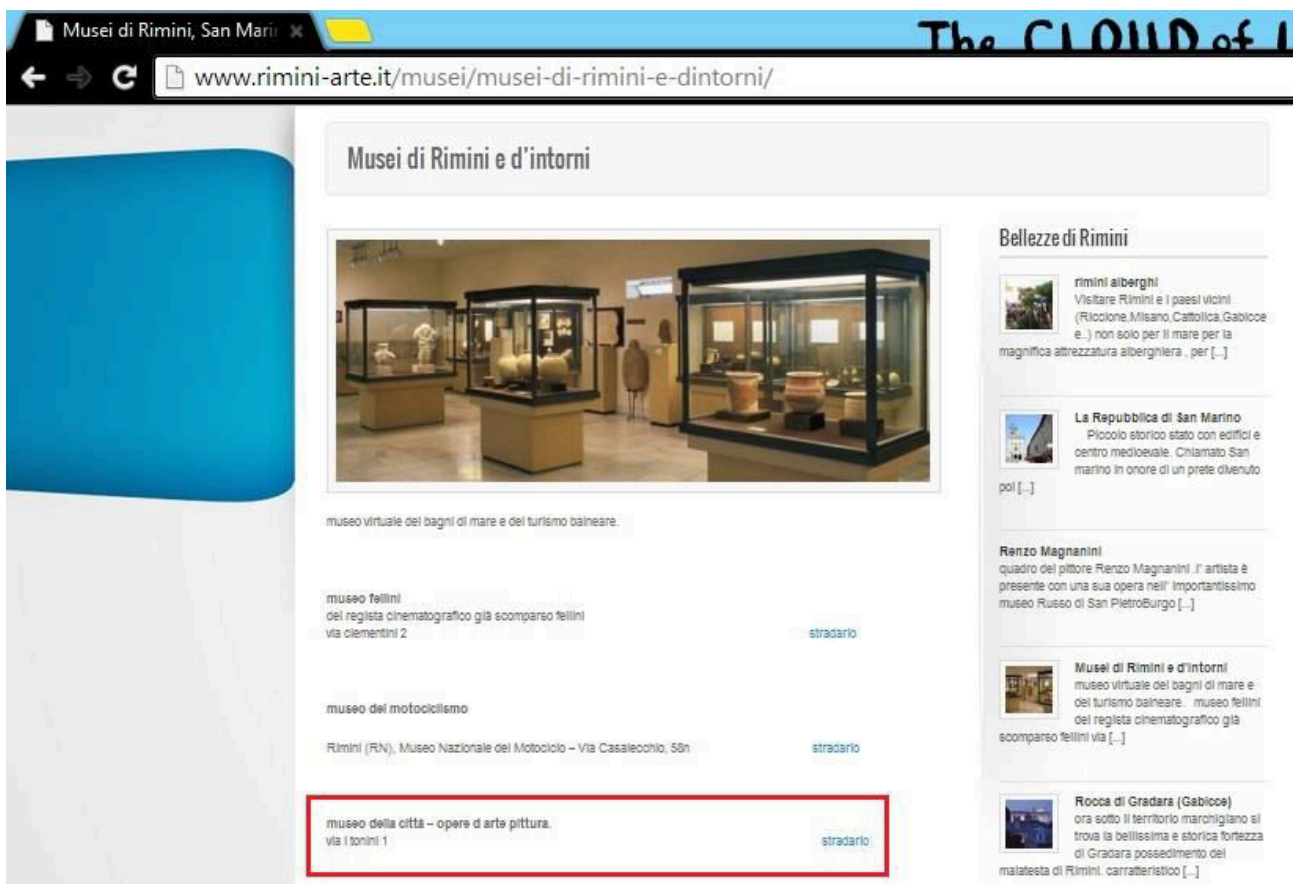
L'obiettivo del potenziamento del sito è quello di attirare più turisti, superando per esempio quelle che sono le alternative proposte dai motori di ricerca quando viene effettuata una query generica su Rimini. Ad esempio, se un utente effettua una ricerca inserendo unicamente come keyword la parola "Rimini", ottiene diversi milioni di risultati, e si è già detto nell'introduzione che il sito del museo non è sicuramente tra questi. Se invece vengono inserite altre parole chiave, raffinando la ricerca allora il sito museale ha più possibilità di apparire agli utenti. Si reputa quindi necessario fare alcune prove per vedere con quali parole chiave ci sono più possibilità di fruizione da parte del museo, al termine di questo capitolo si cercherà di fornire una stima di ROI, nel caso in cui vengano inserite altre parole chiave e modificate quelle attualmente presenti.

1.1 Le parole chiave

Su WooRank.com viene segnalata la cosiddetta "keywords cloud", ovvero un insieme di parole chiave più ricorrenti e con maggiori probabilità di ottenere un buon ranking nei search engine. Le keyword elencate sono: rimini, museo, città, settembre, mostre, musei, data, domus, sguardi

Per queste parole chiave viene anche segnalato il fattore consistenza, giudicato corretto. Ciò significa che le i termini più significativi compaiono frequentemente sia nei titoli, che nelle descrizioni e heading tag. Ad ogni modo, se si effettua su Google una ricerca per la parola "Rimini", al primo posto risulta la pagina Wikipedia della città, seguita immediatamente dai siti

Rimini Turismo¹³³ (prima in lingua inglese e poi in italiano), la pagina del comune, quella della provincia di Rimini, la pagina dell'aeroporto, della riviera¹³⁴ e infine i consigli di viaggio di Wikitravel e TripAdvisor. Del sito museale non vi è traccia fino alla quarta pagina. Non si hanno risultati migliori tentando di raffinare la ricerca inserendo nella query “Rimini arte”. In questo caso il sito museale figura solo in quinta pagina, mentre ai primi posti si trovano il portale turistico Rimini-Arte¹³⁵ e la pagina dedicata agli eventi di Rimini¹³⁶, con riferimento alle mostre riguardanti Federico Fellini e quelle di arte contemporanea presso FAR | fabbrica d'arte di Rimini. In nessuno dei due risultati figurano dei collegamenti diretti al sito dei musei della città. Addirittura, nel portale turistico Rimini-Arte c'è la possibilità di selezionare la voce “Musei”, ma, come riportato dall'immagine, alla voce Museo della Città (peraltro situata in terza posizione dopo il museo Fellini –attualmente chiuso- e quello del motociclismo), non è presente un link di rimando alla pagina, solo un collegamento a Google Maps per mostrarne la localizzazione.



Il ranking del sito museale non migliora nemmeno se si effettuano ricerche con parole come “Rimini cose da vedere” o “Rimini monumenti da visitare”, per cui si aprono in prima posizione le pagine di TripAdvisor e il portale di Rimini Eventi. A questo punto non resta che raffinare di più la

133 http://www.riminiturismo.it/CMS2/main.php?lang_index=1

134 <http://www.riviera.rimini.it>

135 <http://www.rimini-arte.it>

136 http://www.rimini.com/eventi/index/cat_eventi-20

ricerca, con query del tipo “Rimini Museo” e direttamente “Museo della città di Rimini”. Almeno in questi casi, il sito in questione figura subito come primo risultato di ricerca. Alexa.com suggerisce che queste siano proprio le parole chiave principali digitate dagli utenti nei motori di ricerca:

- rimini museo per il 9,64%;
- museo della città di rimini per il 9,36%;
- museo della citta rimini per il 4.88%;
- rimini coop per il 4.77%.

1.2. Analisi degli utenti

Lo studio svolto da Alexa.com torna più utile in questo paragrafo, tuttavia, prima di procedere si vuole fare una piccola segnalazione relativa all'affidabilità di Alexa.com, che non è 100% effettiva. Questo perché i dati raccolti dal team di Alexa dipendono e fanno riferimento agli utenti che hanno installato la loro toolbar. È proprio questo strumento che fornisce i dati per le statistiche in seguito sviluppate. Sostanzialmente le stime riguardanti il sito dei musei di Rimini dovrebbero fare riferimento agli utenti che hanno visualizzato il sito con la toolbar installata, in grado di passare le informazioni al sito di analisi, ma si può immaginare la limitatezza di tali dati. È l'azienda stessa che specifica i propri limiti nel suo website, scrivendo: “Alexa’s traffic estimates are based on a diverse sample of millions of worldwide internet users using thousands of different types of toolbars and add-ons for Google Chrome, Firefox, and Internet Explorer.”¹³⁷ Questo può condurre ad errori nelle stime riportate dall'azienda, stime che cercano di essere corrette, ma ciò è impossibile da risolvere completamente e le stime di traffico possono essere superiori o inferiori al traffico effettivo del sito. Detto questo si procede a riportare quanto rilevato dall'analisi, a partire dal fatto che l'81,5% degli utenti del sito sono italiani. Questo può dipendere dal fatto che le parole chiave siano solo in italiano. È vero che esiste la versione in inglese del sito, ma forse una maggiore pubblicità in circuiti internazionali potrebbe essere necessaria. Senza considerare che è possibile selezionare la voce “english” nel menu della home page, ma ben pochi contenuti sono disponibili in lingua inglese e di quei pochi, si deve sottolineare la loro traduzione con programmi automatici, con conseguenti errori e sviste.

In seguito allo studio della stringa referrer all'interno dei file di log del web server si è anche scoperto che il 56,3% delle visite dirette al sito dei Musei deriva dal sito del comune della città (comune.rimini.it).

137 www.alexa.com/company

How engaged are visitors to museicomunalirimini.it?



L'immagine riportata mostra alcuni dati reperiti da Alexa riguardo i comportamenti degli utenti. La bounce rate dei visitatori del sito è del 50,00% (con una diminuzione recente del 5,00%). Questo valore costituisce la percentuale di visite che in seguito alla visualizzazione di una pagina hanno poi abbandonato il sito. Il dato è da valutare tenendo conto di alcuni fattori, come per esempio il fatto che spesso gli utenti accedano al sito del museo per motivi logistici, quali orari o giorni di chiusura, e una volta soddisfatta la richiesta non hanno più bisogno o interesse nel sito. Il secondo dato riportato riguarda il numero di pagine visualizzate quotidianamente per visitatore, che è di 3.70 pagine (con un aumento recente del 40.00%). Il terzo elemento invece raffigura il tempo medio che un utente trascorre sul sito, che è di 2:12.

2. Analisi interna

2.1. Il nome del dominio

La scelta che è stata fatta per il sito Web dei musei è: www.museicomunalirimini.it

Questo significa che il nome dell'host è "museicomunalirimini", mentre il TLD è .it. Si è detto nel secondo capitolo delle norme di scelta di un buon nome di dominio; in questo caso, pur trattandosi di un nome abbastanza lungo, il titolo scelto risulta chiaro ed informativo. Fa riferimento sia alla città di Rimini, chiaro elemento di localizzazione geografica, sia alla gestione museale, che è comunale, oltre a rispecchiare la tipologia d'esercizio, contiene alcune keyword, il che è molto importante. Secondariamente, si è già scritto di come è consigliabile acquisire varie estensioni del nome del dominio per evitare che altri ne prendano i diritti e possano poi farne uso; nel caso del museo Woorank suggerisce queste come disponibili:

- museicomunalirimini.com;
- museicomunalirimini.net;
- museicomunalirimini.org;
- museicomunalirimini.info;
- museicomunalirimini.biz;

Se venissero acquisiti dai Musei Comunali si potrebbe fare in modo che rimandassero alla stessa

pagina, attirando una quantità di traffico ancora maggiore.

2.2. Keyword density

Sfruttando la presenza di alcuni tool disponibili online, quali Keyword Density Calculator¹³⁸, è possibile visualizzare le parole chiave più comuni e il loro indice KD.

Analisi pagina: <http://www.museicomunalirimini.it/>

Parole: 57, Quantità: 179			Frase: 30, Quantità: 69			Frase: 23, Quantità: 49		
Parola	Quantità	Densità	2 Parole	Quantità	Densità	3 Parole	Quantità	Densità
2013	11	6.15%	della città	5	7.25%	museo della città	5	10.2%
di	10	5.59%	museo della	5	7.25%	della città dal	2	4.08%
rimini	8	4.47%	settembre 2013	3	4.35%	città dal 31	2	4.08%
museo	7	3.91%	il museo	3	4.35%	dal 31 agosto	2	4.08%
settembre	6	3.35%	di rimini	3	4.35%	far museo della	2	4.08%
città	6	3.35%	immagine far	2	2.9%	immagine far museo	2	4.08%
al	5	2.79%	far museo	2	2.9%	rimini foto di	2	4.08%
della	5	2.79%	dal 31	2	2.9%	foto di settembre	2	4.08%
la	5	2.79%	dell'immagine	2	2.9%	galleria dell'immagine	2	4.08%
dal	4	2.23%	foto di	2	2.9%	dell'immagine far	2	4.08%
musei	4	2.23%	per la	2	2.9%	agosto al 29	2	4.08%
29	4	2.23%	rimini foto	2	2.9%	29 settembre 2013	2	4.08%
del	4	2.23%	31 agosto	2	2.9%	percorsi letture mostre	2	4.08%
per	4	2.23%	di settembre	2	2.9%	letture mostre spettacoli	2	4.08%
31	3	1.68%	galleria dell	2	2.9%	mostre spettacoli feste	2	4.08%
mese	3	1.68%	al 29	2	2.9%	spettacoli feste giochi	2	4.08%
data	3	1.68%	letture mostre	2	2.9%	laboratori percorsi letture	2	4.08%
percorsi	3	1.68%	percorsi letture	2	2.9%	conferenze laboratori percorsi	2	4.08%
il	3	1.68%	mostre spettacoli	2	2.9%	mese delle famiglie	2	4.08%
mostre	3	1.68%	spettacoli feste	2	2.9%	delle famiglie 2013	2	4.08%
appuntamenti	3	1.68%	feste giochi	2	2.9%	convegni conferenze laboratori	2	4.08%
18	3	1.68%	laboratori percorsi	2	2.9%	al 29 settembre	2	4.08%

Per quanto riguarda le keyword short tail, le più diffuse sono:

Parola	Quantità	Densità
2013	11	6.15%
rimini	8	4.47%
museo	7	3.91%
settembre	6	3.35%
città	6	3.35%
mostre	3	1.68%
percorsi	3	1.68%
far museo	2	2.9%
convegni	2	1.12%
laboratori	2	1.12%
giochi	2	1.12%
famiglie	2	1.12%

Mentre per quanto riguarda le keyword long tail, le più diffuse sono:

3 Parole	Quantità	Densità
museo della città	5	10.2%
letture mostre spettacoli	2	4.08%
mostre spettacoli feste	2	4.08%
spettacoli feste giochi	2	4.08%

¹³⁸ tool.motoricerca.info/density-keyword.phtml

laboratori percorsi letture	2	4.08%
immagine far museo	2	4.08%
galleria dell immagine	2	4.08%
conferenze laboratori percorsi	2	4.08%
mese delle famiglie	2	4.08%

La keyword density dovrebbe avere un valore compreso circa tra il 3 e il 5%, nel caso questo valore venga superato si rischia sempre il keyword stuffing. Questi valori si intendono per le parole chiave più rilevanti ai fini della propria campagna SEO; in queste due tabelle molte parole non risultano essere focalizzate. Contemporaneamente la quantità di alcune keyword essenziali risulta essere scarsa, per esempio: mostre, percorsi, opere d'arte, visite, educazione, creatività, esibizioni, storia, romani, monete, ecc. per non citare i nomi di musei e sedi, quali: Museo degli Sguardi, FAR o la Galleria dell'immagine.

2.3. Meta tag

Tramite lo strumento "sorgente" presente su Google Chrome è possibile visualizzare il codice HTML che è stato inserito per la home page del sito dei Musei Comunali. Nei prossimi paragrafi si intendono analizzare nel dettaglio tutte le porzioni del codice, valutandone l'efficacia e le eventuali mancanze, ma per ora lo si riporta nella sua interezza:

```

        <head>
        <title>Musei di Rimini</title>
        <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
        <meta name="Author" content="Officine Digitali" />
        <meta name="Contact" content="info@officinedigitali.it" />
        <meta name="description" content="Obiettivo principe del sito dedicato ai musei di Rimini è quello di 'illustrare' compiutamente la storia e lo svolgimento culturale della città e del suo territorio. Il sito contiene infatti informazioni e immagini afferenti al Museo della Città e alla domus del chirurgo, nonché al Museo degli Sguardi. Inoltre le informazioni contenute nel sito vengono automaticamente trasferite sul portale europeo dei musei (www.euromuse.net), promuovendo gli stessi in una dimensione internazionale. Il sito ha così una versione italiana e una inglese." />
        <meta name="keywords" content="museo, musei, domus, sguardi" />
        <link rel="stylesheet" href="css/stile.css" type="text/css" media="screen" />
        <!-- <link rel="stylesheet" href="css/stile.php?h=1" /> -->
        <link rel="stylesheet" href="css/stile.php" type="text/css" />
        <link rel="stylesheet" href="css/print_home.css" type="text/css" media="print" />
        <script type="text/javascript" src="js/AC_RunActiveContent.js"></script>
        <script type="text/javascript" src="js/odSlide.js"></script>
        </head>139

```

139 view-source: <http://www.museicomunalirimini.it/>

2.4. Title tag

In passato il title tag era considerato sostanzialmente come il fattore di ranking on-page più rilevante in assoluto; attualmente i tempi (e le tecnologie) sono cambiati e la sua importanza è stata ridimensionata. Si sono già descritte le migliori norme di compilazione e la possibilità di inserire al suo interno oltre le parole chiave anche i nomi delle aziende o le città di riferimento. Questo tag viene mostrato come “headline” nella SERP, e, assieme alla descrizione della pagina, costituisce una grossa determinante nella scelta che il visitatore si trova ad effettuare dopo aver effettuato la ricerca. Il titolo viene anche ripetuto nel margine superiore della scheda del browser, è il nome automatico che viene dato alla pagina quando essa viene inserita tra i preferiti, ma non appare mai veramente “nella pagina”.

Nel caso in analisi, il codice inserito è: <title>Musei di Rimini</title>

La lunghezza risulta corretta, si tratta di 15 caratteri, per cui rientra ampiamente nei limiti, che erano di massimo 70. È importante che all'interno del titolo siano contenute le parole chiave principali e che per ogni pagina ci sia un titolo unico. Nel sito museale ogni pagina risulta avere un titolo specifico, se dal menu della home page si seleziona la voce “musei”, il title tag della nuova pagina che si apre è “Musei di Rimini – Musei”. A questo proposito, sarebbe meglio sostituire il trattino orizzontale con un pipe, in questo modo: “Musei di Rimini | Musei”. Lo stesso vale per la pagina intitolata “Musei di Rimini - Orari/Biglietti/Prenotazioni” si consiglia di cambiare in “Musei di Rimini | Orari | Biglietti | Prenotazioni”.

Nel complesso la voce title tag risulta corretta, anche se WooRank suggerisce la possibilità di inserire un numero maggiore di parole chiave, specifiche per ogni pagina, per esempio risultano assenti keywords come: guida, visita guidata, ingressi gratuiti, ecc.

2.5. Meta description

La descrizione della pagina Web viene utilizzata sia dai motori di ricerca che dalle directories. Essa può influenzare i ranking ed essere visualizzabile nella SERP, sempre elemento incidente sulla scelta da parte del visitatore, se cliccare o meno. La descrizione inserita nel sito è la seguente:

“Obiettivo principe del sito dedicato ai musei di Rimini è quello di 'illustrare' compiutamente la storia e lo svolgimento culturale della città e del suo territorio. Il sito contiene infatti informazioni e immagini afferenti al Museo della Città e alla domus del chirurgo, nonché al Museo degli Sguardi. Inoltre le informazioni contenute nel sito vengono automaticamente trasferite sul portale europeo

dei musei (www.euromuse.net), promuovendo gli stessi in una dimensione internazionale. Il sito ha così una versione italiana e una inglese.¹⁴⁰

La lunghezza è di 541 caratteri, quindi nettamente superiore ai 160 caratteri, spazi inclusi, che si erano detti soglia massima. Tra le varie conseguenze di questa lunghezza ce n'è una esteticamente più rilevante, ovvero che nelle SERP la descrizione del sito appare tagliata:

Musei di Rimini

www.museicomunalirimini.it/ ▼

Obiettivo principe del sito dedicato ai **musei di Rimini** è quello di 'illustrare' compiutamente la storia e lo svolgimento culturale della città e del suo territorio. Il sito ...



Museo della Città

Museo della Città. Banner museo della città. MUSEO DELLA ...

Domus del chirurgo

Domus del chirurgo. Attenzione S necessario installare o ...

Orari/Biglietti/Prenotazioni

Museo della Città e Domus del Chirurgo orari dal 1 settembre ...

Sponsor

Awiso ricerca sponsor. Contratto di sponsorizzazione. L ...

Dove siamo / Come arrivare ...

Sei in: Come visitare i Musei » Dove siamo / Come arrivare ...

Orari estivi dei Musei Comunali

Dal 1 giugno al 31 agosto 2013 entrano in vigore gli orari estivi ...

[Altri risultati in museicomunalirimini.it »](#)

La descrizione influenza come le pagine web vengono descritte e mostrate nei motori di ricerca; una buona descrizione agisce come una potente forma di pubblicità organica e incoraggia l'utente a cliccare il proprio sito. Deve essere esplicita e contenere le parole chiave più importanti. Ogni pagina dovrebbe avere un'unica descrizione relativa agli esatti contenuti, ma nelle varie pagine che compongono il sito dei musei di Rimini la descrizione è la stessa, ripetuta su tutte le pagine. Questo non aiuta né gli utenti né i motori di ricerca che fanno più fatica a capire i contenuti di ogni documento. Non ha senso che pagine diverse come quella relativa all'accessibilità ai musei, quella intitolata "Musei di Rimini - Notizie e Musei" e quella "Perché Riminimusei" abbiano la stessa descrizione della homepage.

2.6. Rich snippet

Il sito dei musei di Rimini non utilizza microformati, la mancanza di questo markup impedisce a Google di migliorare la presentazione del sito all'interno dei risultati di ricerca. Per questo motivo non è possibile visualizzare rich snippet o recensioni del sito. Tra l'altro viene segnalata da WooRank anche l'assenza di dublin core, ovvero un set di metadati usati per descrivere i contenuti di un sito. Ad esempio possono essere utili per organizzare il layout visualizzato dai motori di

140 <http://www.museicomunalirimini.it>

ricerca interni e non disturbano il codice della pagina.

2.7. Keywords meta tag e language meta tag

Nel secondo capitolo si è scritto ampiamente di questo controverso tag, in questo caso lo sviluppatore del sito museale ha scelto di implementarlo, e la stringa è la seguente:

```
<meta name="keywords" content="museo, musei, domus, sguardi" />
```

Il numero massimo di keyword inseribili è di sei parole, per cui quanto inserito risulta corretto. L'unica cosa che si può aggiungere è che, nel momento in cui si procede ad aumentare il numero di pagine che compongono il sito, questo tag può anche essere tralasciato. Per quanto riguarda il tag relativo al linguaggio, i sistemi analitici l'hanno identificato e verificato come italiano.

2.8. Robots meta tag

Questi tag sono utili perché consentono di fornire direzioni ben precise ai crawler dei motori di ricerca. Come segnalato da WooRank e altri siti d'analisi, il website dei musei di Rimini non presenta alcun robots meta tag. Si ricordano alcune importanti funzioni di questi file, per cui se, per esempio, si ha il proprio dominio inserito nelle directories di Yahoo! o DMOZ e si vuole evitare che la descrizione della pagina venga presa di lì si possono usare tag come CONTENT="NOODP" e CONTENT="NOYDIR", così facendo viene inserita quella desiderata e segnalata nei meta tag. I file robots.txt sono fortemente raccomandati all'interno del dominio, o si rischia di generare la pagina automatica "404 errors" quando i search engine cercano, e non trovano, il dominio del sito del Museo. L'assenza di file è in grado di generare anche altri errori nel proprio apache log, quando gli spider ne richiedono uno. Allo stesso tempo un'altra importante funzione dei file robots e quella di tenere lontani i motori di ricerca dalle aree del sito che non si desidera che vengano scansionate e indicizzate. Infine, i robots.txt sono necessari come referenza per la localizzazione del file sitemap.xml.

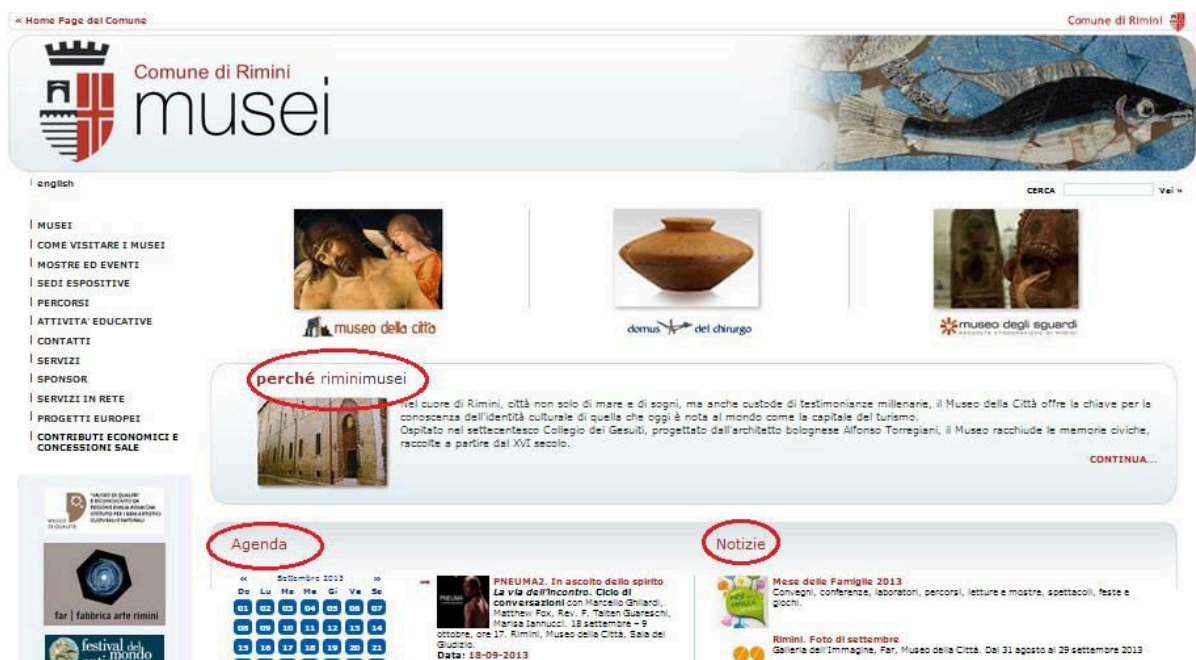
La mancanza di questi tag viene indicata come un'emergenza da numerosi siti, che raccomandano di aggiungere un file robot subito, anche utilizzando dei generatori gratuiti, come questo Robots.txt File Generator¹⁴¹. Oltre alla protezione dei contenuti, i file robots.txt possono far risparmiare sulla larghezza di banda e diminuire il carico sul server.

2.9. Heading tag

Il tag H1 dovrebbe avere sempre gli stessi contenuti del title tag, dato che costituisce la testata della pagina. Questo tag è importante perché rende i contenuti di pagina subito evidenti e chiari al

141 <http://www.seoautomatic.com/robotstxt-file-generator>

visitatore; nel caso del sito dei musei della città di Rimini la testata utilizzata è: Musei di Rimini. Per renderla ancora più utile, si potrebbero utilizzare dei text link verso le altre pagine più rilevanti. Vengono utilizzati anche i tag H2 che contribuiscono a marcare l'argomento di una pagina, aiutandone l'identificazione per il ranking. Come illustrato in figura, nel sito del museo sono presenti diversi tag H2: perché Rimini musei, agenda, notizie. Se si hanno pagine sufficientemente lunghe si possono avere più testate di livelli diversi e l'utilizzo dei vari header tag, come visto nel secondo capitolo, può aiutare a definirne la priorità. Bisogna fare attenzione ad utilizzare un solo H1 tag, ma si possono usare più sub-testate in una stessa pagina. Il sito dei musei fa uso di diverse testate ed è stato sottolineato che si potrebbero utilizzare anche i testi in grassetto invece dei tag H3, H4, ecc., non sembra che ci siano particolari vantaggi ad usare uno piuttosto che l'altro.



2.10. Ottimizzazione della pagina

Gran parte dalle analisi diagnostiche svolte dai vari siti si concentrano sulla struttura della home page del sito e su come possa essere ottimizzata. Sono emersi fattori positivi, come la totale assenza di animazioni flash, ma anche molto negativi come i lunghi tempi di caricamento necessari per poter accedere alla pagina.

2.10.1. Animazioni Flash

Il sito del museo non contiene né animazioni flash, né frame. Questo è un fattore molto positivo, entrambi gli elementi possono causare grossi problemi alla pagina Web che li contiene,

principalmente perché i motori di ricerca non sono in grado di scansionarne e indicizzarne i contenuti. Questo succede perché ogni frame è una pagina a parte, quindi richiede tempi di caricamento maggiori, oltre all'uso di un raccogliatore dei vari frame (frame set) che porta ad ulteriori complicazioni. Se si ha necessità di inserire dei frame all'interno di una webpage allora si consiglia l'utilizzo di un <NoFrames> tag.

2.10.2. Percentuale testo/codice

Woorank ha rilevato che nella homepage del sito del museo comunale la percentuale di testo rispetto al codice HTML è del 4.59%. I punteggi ideali in questo campo dovrebbero essere compresi tra un minimo di 25% e un massimo di 70%. Per percentuali inferiori, come nel caso del sito del museo, allora sarebbe il caso di inserire più text content, rispetto alla parte codificata. In caso di punteggi superiori al 70%, allora gli sviluppatori del sito rischiano di far passare la pagina come spam, agli occhi dei motori di ricerca. Come linea generale però si può sostenere che, fintanto che non si esageri e i contenuti espressi siano rilevanti, più testo è contenuto dalla pagina, meglio è.

2.10.3. Dimensioni della pagina

Per quanto riguarda le dimensioni della pagina, la media diffusa all'interno del World Wide Web è di circa 320KB, mentre la homepage dei musei riminesi è di 420.2 KB. Questo significa che la pagina necessita di essere alleggerita, principalmente per evitare i lunghi tempi di caricamento; bisognerebbe almeno scendere sotto la soglia dei 300 KB. I due principali fattori che incidono in questo caso potrebbero essere le dimensioni delle immagini e i file JavaScript. Nel prossimo paragrafo si approfondisce l'aspetto relativo alle immagini, per quanto riguarda JavaScript invece, si tratta di “un linguaggio di scripting orientato agli oggetti, comunemente usato nei siti web. Può essere adoperato in concomitanza con altre tecnologie come Ajax e Silverlight”¹⁴². Più semplicemente, si tratta di un linguaggio che consente ai siti Web di essere più interattivi e possedere contenuti unici. Il sito fortunatamente possiede pochi contenuti JavaScript, ad ogni modo le normali precauzioni da prendere sono l'utilizzo di file JS esterni alla pagina, sfruttare il tag <noscript> in grado di fornire del testo alternativo ai motori di ricerca e infine utilizzare la funzione google.load le librerie JavaScript o Ajax necessarie. Se c'è necessità di comprimere i file JavaScript, due programmi utili sono Closure Compiler e YUI Compressor, entrambi gratuiti e consentono di scegliere la tipologia di compressione preferita. Sia WooRank che SEOAutomatic hanno rilevato i tempi di caricamento del sito come troppo lunghi. In seguito ad un page speed test

142 Maltraversi, 2012, pag. 247

effettuato tramite lo strumento d'analisi Pingdom.com, di cui si sono riportati i risultati in figura, si è visto che i tempi di caricamento sono dell'ordine di 3.43 secondi.



I lunghi tempi di caricamento costituiscono spesso il principale motivo per cui la maggioranza degli utenti abbandona il sito web dopo poco. Proprio per questo motivo Google fornisce uno strumento online che mostra ai web developers come riuscire ad accelerare i propri siti, chiamato Page Speed Online¹⁴³. Spesso le tempistiche di caricamento possono dipendere dal server, o da altri problemi di rete, ma anche da una cattiva ottimizzazione del codice. Esistono poi degli strumenti di website monitoring che consentono di monitorare il proprio server e di ricevere segnalazioni tramite SMS quando i tempi sono troppo lunghi, come per esempio monitor.us o siteuptime.com. Ad ogni modo, il server utilizzato dal sito del museo non impiega *cached pages*. L'utilizzo della cache (HTTP caching), permette un caricamento rapido delle singole pagine in seguito ad una prima visita. Sostanzialmente si tratta di una memoria temporanea, che salva alcuni dati automaticamente, in modo da facilitarne il recupero su richiesta. Applicandola ad un server proxy riesce a ridurre il carico di richieste, contribuendo al miglioramento delle prestazioni del sistema. Ad esempio si risparmia sulla larghezza di banda e si velocizzano i tempi di caricamento, anche mostrando versioni statiche delle proprie pagine dinamiche.

L'analisi svolta su Search Commander ha portato alcuni risultati in merito al peso della pagina. La dimensione del codice HTML non compresso è di 51.55 KB. Questo peso non include i file JavaScript esterni, le immagini e i file CSS. Un valore ottimale dovrebbe essere sotto i 25KB per cui, una riduzione del peso è necessaria. Per quanto riguarda lo status di compressione HTML, le pagine compresse chiaramente caricano più velocemente; il sito del museo fa uso di Gzip¹⁴⁴, un'applicazione che consente di ridurre il peso di JavaScript e CSS. La percentuale di compressione esercitata dalla pagina web è del 13%, per cui il risultato è considerato come positivo.

Vi sono alcune note positive rilevate dalla diagnosi della pagina web, ovvero anzitutto non sono

143 <http://code.google.com/speed/page-speed>

144 Una guida rapida è reperibile a questo indirizzo: <http://www.regole-seo.com/gzip-compression>

utilizzate tabelle nidificate all'interno del sito; questo è un bene poiché più le pagine sono articolate, maggiori sono i tempi di riproduzione del layout da parte del browser, che prima di poter caricare l'intera pagina deve aspettare di essere arrivato in fondo alla tabella. Per quanto riguarda l'analisi dei file CSS (Cascading Style Sheets), essi costituiscono un linguaggio impiegato dagli sviluppatori di siti Web per definire la formattazione di documenti HTML, XHTML e XML, ovvero l'aspetto che le pagine avranno agli occhi dei visitatori. La homepage del sito dei musei di Rimini contiene pochi CSS, il che ne riduce la pesantezza, ma se un sito ne avesse in quantità maggiori esistono programmi come CSS Optimizer o Clean CSS, che consentono di comprimerli. Uno dei file presenti, giudicato positivo da WooRank, è il "print-friendly CSS". Questo speciale foglio di stile consente all'utente di stampare in maniera diretta le pagine del sito, automaticamente modificandone il design. In questo modo non vengono riprodotti banner di interfaccia o immagini, consentendo sia di evitare stampe di elementi che prendono spazio inutilmente, sia di risparmiare inchiostro della stampante. Uno dei file CSS utilizzati, giudicato negativo da WooRank, è "inline styles". Si tratta di file CSS che sono applicati ad ogni singolo elemento utilizzando l'attributo "style". Il lato negativo del loro uso è che devono essere ripetuti per ogni frase, per cui se per esempio si desidera che tutti i paragrafi componenti una pagina abbiano lo stesso font "Arial", allora bisogna aggiungere un inline style per ogni <p> tag del documento (il tag che ne influenza il carattere). Questo costituisce sia un lavoro in più per il webmaster, sia tempi di caricamento maggiori per l'utente che deve caricare la pagina.

Altra feature giudicata positiva dai siti di analisi SEO è la presenza di breadcrumbs nelle pagine, che consentono al visitatore di sapere costantemente la sua posizione rispetto alle pagine del sito.



2.11. Ottimizzazione delle immagini

2.11.1 Titolo delle immagini

Per ottimizzare al meglio le immagini presenti sul sito web bisogna rispettare tre fattori fondamentali. Il primo riguarda i nomi delle immagini, che devono essere brevi e richiamare l'argomento della foto senza lasciare spazio a dubbi. Questo perché alcuni utenti potrebbero raggiungere il sito Web tramite l'utilizzo di Google Image o Google Universal Search e quindi avere nomi appropriati può solo che agevolare il processo di ricerca. A questo proposito le trenta immagini contenute nella homepage del sito dei musei di Rimini presentano quasi tutte titoli appropriati e descrittivi, anche per respingere ogni rischio di apparire come spam.

Alcuni titoli sono i seguenti:

- Logo Portale Musei Comune di Rimini



- Festival del Mondo Antico



- Teatro Ermete Novelli
- Romagna Visit Card
- Banner Museo della Città

2.11.2. Attributo Alt

Il secondo fattore che deve essere verificato riguarda l'attributo alt, che deve essere sia presente che rispecchiare anch'esso il contenuto dell'immagine. Se per caso ci fossero dei problemi alla visualizzazione dell'immagine, l'attributo verrebbe visualizzato consentendo comunque agli utenti di capire di cosa si tratti all'interno della pagina. Tutte le immagini contenute nella homepage hanno l'attributo alt completato, nessuno è stato lasciato vuoto. Questo significa che Google è in grado di mostrare le immagini all'utente che effettua una ricerca su Google Images.

2.11.3. Tag width e tag height

Il terzo fattore riguarda i tag width e height, la cui presenza è essenziale per il browser che non dovrà calcolarne le dimensioni per visualizzare le immagini. Oltre a questo, l'utilizzo dei tag consente al browser di caricare la pagina senza dover aspettare di effettuare il download delle immagini ivi contenute. Sostanzialmente, se si scegliesse di non implementare i due tag, allora l'intera pagina si bloccherebbe durante il rendering dell'immagine, rallentando i tempi di caricamento e conseguentemente riducendo la frequenza e profondità di scansione della pagina da parte dei crawler dei motori di ricerca. La SEO review svolta da Search Commander mostra che alcune immagini sono prive di questi due attributi. Questo significa che la pagina non viene mostrata all'utente fino a che tutte le immagini siano state scaricate, con conseguenti rallentamenti. Un elemento che può aiutare velocizzare i tempi di caricamento è l'utilizzo di "image expires" nell'intestazione delle immagini. Se si sceglie di applicarlo, allora i browser memorizzano le immagini tramite cache per un determinato periodo, facilitando la visualizzazione agli utenti che ritornano al sito.

2.12. Scrittura delle URL

Lo studio condotto su WooRank certifica che il fattore URL Rewrite è adeguato e questo perché gli indirizzi URL sono puliti, privi di query strings. L'utilizzo di URL pulite oltre ad essere seo-friendly è anche importante ai fini dell'usability. L'unico elemento che è stato segnalato su cui bisogna fare attenzione riguarda la presenza di underscore nelle URL. Nella homepage del sito (o negli URL presenti al suo interno) sono stati trovati indirizzi con underscore, sarebbe meglio utilizzare hyphens (ovvero dei trattini), perché questi ultimi vengono percepiti da Google come dei separatori di parole, a differenza dell'underscore. Un esempio pratico, alla voce "come visitare i musei", l'URL invece che essere:

www.museicomunalirimini.it/visitare_musei/

dovrebbe essere:

Sempre a proposito del fattore usability, la pagina web non presenta una favicon; essa è costituita dal logo che appare normalmente nella barra indirizzo, di fianco al proprio URL. Il suo utilizzo sarebbe molto utile per aumentare la visibilità del proprio brand e per facilitare il ritorno degli utenti al sito dei musei quando essi lo cercano all'interno dei propri preferiti. Il logo dei musei di rimini potrebbe essere inserito sia in posizione di favicon, ma anche in altre parti del sito in modo da poter essere facilmente riconoscibile. Oltre all'effetto sul ranking, effettuare una scelta simile porterebbe risultati anche da un punto di vista di ricavi di gestione straordinaria: se il logo viene diffuso e diventa distintivo, si possono creare gadget e merci branded, vendibili nel bookshop. Per poter mostrare una favicon, alcuni siti utilizzano semplicemente il file .ico, e in questo modo sostanzialmente sono i browser che autonomamente rintracciano l'immagine. Alcuni generatori di Favicon sono rintracciabili a questi due indirizzi:

1. <http://tools.dynamicdrive.com/favicon/#.UkmhG9JSgQg>;
2. <http://www.favicongenerator.com>;

Per quanto riguarda la gestione dei link, si comincia dell'anchor text. Si è già parlato della sua rilevanza soprattutto in seguito all'aggiornamento di Google Penguin dell'aprile del 2012, attualmente è fondamentale che il testo inserito nei link sia rilevante e collegato alla pagina che poi si va ad aprire. Quando non è possibile creare dei link con un buon testo, o i collegamenti rischiano di apparire troppo artificiali, si può utilizzare il testo intorno ad essi per invogliare gli utenti, evitando frasi tipo "click here". Alcuni esempi di link interni:

- mostre ed eventi; www.museicomunalirimini.it/mostre_eventi/
- sedi espositive; www.museicomunalirimini.it/sedi_espositive/
- attività educative; www.museicomunalirimini.it/attivita_educative/

2.13. Google Sitelink

I Google site link sono dei link aggiuntivi che Google genera per aiutare gli utenti a navigare meglio su un determinato sito. Sostanzialmente guidano i visitatori tra i vari contenuti direttamente dalla SERP.



Il sito dei musei riminesi presenta questi sitelink e questo è molto positivo, in quanto non è ben chiaro cosa bisogna fare per riceverli, ma pare che vengano generati automaticamente quando:

- il sito ha un buon traffico (almeno 25 visite giornaliere);
- quando a parere di Google i link presenti siano di buona utilità;
- quando il sito ha una buona struttura e navigabilità e, grazie a diversi link in ingresso di buona qualità, gode anche di una discreta popolarità;

I sitelink possono essere gestiti da Google Webmaster Tools > Configurazione Sito¹⁴⁵.

2.14. Creazione di una sitemap

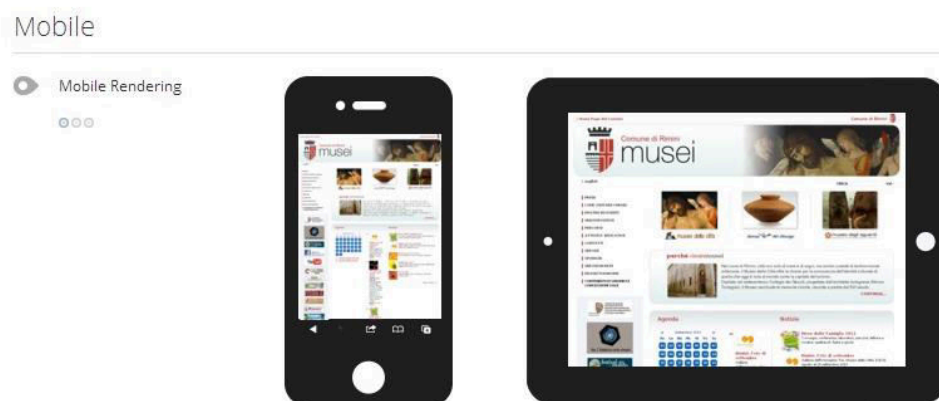
L'utilizzo di una sitemap in formato XML porta il motore di ricerca a capire esattamente dove siano localizzate tutte le pagine del sito Web e la loro rilevanza secondo il parere del webmaster. Il sito dei musei della città è sprovvisto di questo file, e ciò potrebbe significare che o non si è fornita a Google la propria sitemap o che essa sia nominata in maniera diversa. Ad ogni modo la sitemap elenca tutti gli URL disponibili alla scansione e può includere informazioni aggiuntive quali gli ultimi aggiornamenti del sito, la frequenza con cui vengono implementati dei cambiamenti e la rilevanza delle URL. Questo permette ai crawler di scansionare e indicizzare in maniera più efficace e efficiente. Tutti gli strumenti di analisi SEO hanno consigliato di generare una sitemap in formato XML e di inviarla sia a Google webmaster Tools che a Bing Webmasters Tools. È anche buona norma specificare la posizione della sitemap nei propri file robots.txt.

2.15. SEO mobile

Il numero di persone che al giorno d'oggi fanno uso di smartphone, tablet e simili è incredibilmente alto, più del 75% dei consumatori online. Questi sviluppi tecnologici sono in continua evoluzione, costringendo i proprietari di siti Web a doversi sempre aggiornare e a rendere accessibili i propri

145 Maltraversi, 2012, pp.250-252

contenuti agli apparati mobile. Diversi search engine si sono attrezzati di fronte a queste nuove necessità e hanno reso possibile l'indicizzazione di siti creati esclusivamente per l'accesso da smartphone¹⁴⁶. I problemi che i webmaster si trovano a dover affrontare riguardano la gestione tecnica e la fornitura di contenuti diversi in base alla piattaforma utilizzata dall'utente. A questo proposito si possono utilizzare fogli di stile in modo da identificare la presenza o meno di un mobile e quindi successivamente tramite redirect si spinge l'utente verso il contenuto adatto. Alternativamente si può fornire lo stesso URL per versioni sia desktop che mobile, con un cambio di formato in base allo user agent. Il problema per molti riguarda il rischio di cloaking; si era già spiegato nel primo capitolo di questa tecnica black hat bandita da tutti i motori di ricerca, per cui un webmaster diversifica i contenuti del sito in base a chi effettua l'accesso, se uno spider o un utente umano. Ad ogni modo, come si è detto, alcuni motori di ricerca impiegano uno spider differente nella versione mobile, pertanto la possibilità di venir notificati per cloaking è molto remota. Nel caso del sito del museo, il tempo di caricamento da un supporto mobile è eccessivamente lungo, superiore ai cinque secondi. Quest'attesa comporta la perdita di più del 74% dei propri utenti. Si notificano alcuni tool gratuiti che consentono di implementare versioni mobile, come mobify.me, wirenode.com e onbile.com. Successivamente tramite Google Mobilizer è possibile testare l'aspetto del sito, se raggiunto da un dispositivo mobile. Attualmente l'interfaccia del sito dei musei di Rimini sarebbe questa (immagine fornita da WooRank):



3. Analisi esterna

3.1. Link popularity

Nell'ambito in cui si sta trattando ci sono tre fattori che influenzano definitivamente se una campagna SEO stia avendo i risultati desiderati o meno. Il primo è la popolarità, ovvero quando il

146 Ad esempio Google utilizza l'agente Googlebot-Mobile per identificare i siti che offrono contenuti mobile-friendly

sito è noto all'interno del Web, sia agli occhi dei motori di ricerca, che quelli degli utenti. Secondariamente c'è la reputazione di cui gode il sito, ovvero se siti simili ne parlano, come ne parlano, se lo appoggiano con link in entrata e condivisioni di contenuto. Al terzo posto (ma non per importanza) si parla di autorevolezza, i contenuti diffusi devono essere fondati e non solo volti all'attrazione o a incuriosire¹⁴⁷. Per misurare questi valori si possono utilizzare vari sistemi, primo fra tutti il PageRank della pagina. Onde visualizzare la popolarità del sito, è stato scaricato un Plugin per Chrome, chiamato ChromeSEO¹⁴⁸, che riporta il PageRank della pagina, come illustrato in figura. Il valore assegnato al website dei musei riminesi è di quattro su un massimo di dieci.



Anche Alexa Rank è un sistema di valutazione in grado di valutare circa 10 milioni di siti nel mondo, anche se si è già scritto in precedenza della non totale affidabilità dei risultati. Alexa ha valutato la popolarità del sito dei musei decretandone un global rank di 3,089,996, mentre quello italiano è 167,887. Per capire meglio cosa significhino questi dati, si spiega che se il valore stimato da Alexa equivale ad 1 allora quel sito gode di massima autorevolezza e popolarità.



147 Maltraversi, 2012, pag. 306

148 <https://chrome.google.com/webstore/detail/seo-for-chrome>

Altro sintomo della popolarità di un website è dato dal numero di pagine indicizzate su Google, che nel caso dei musei riminesi è di 3130. Questo numero indica il numero di pagine di cui i motori di ricerca sono a conoscenza per un determinato sito; conseguentemente maggiore è questo numero, maggiori sono le possibilità di essere trovati da utenti. A questo proposito le pagine che sono risultate essere più popolari (oltre alla homepage) sono:

- Musei di Rimini – Accessibilità, riguardante le informazioni logistiche quali orari o contatti;
- Musei di Rimini – Notizie; contente tutti gli ultimi aggiornamenti su mostre, convegni, ..;
- Musei di Rimini – Perché Rimini Musei; pagina che spiega la storia e struttura del museo della città;

3.2. Link building

3.2.1. Scambio e integrazione di link

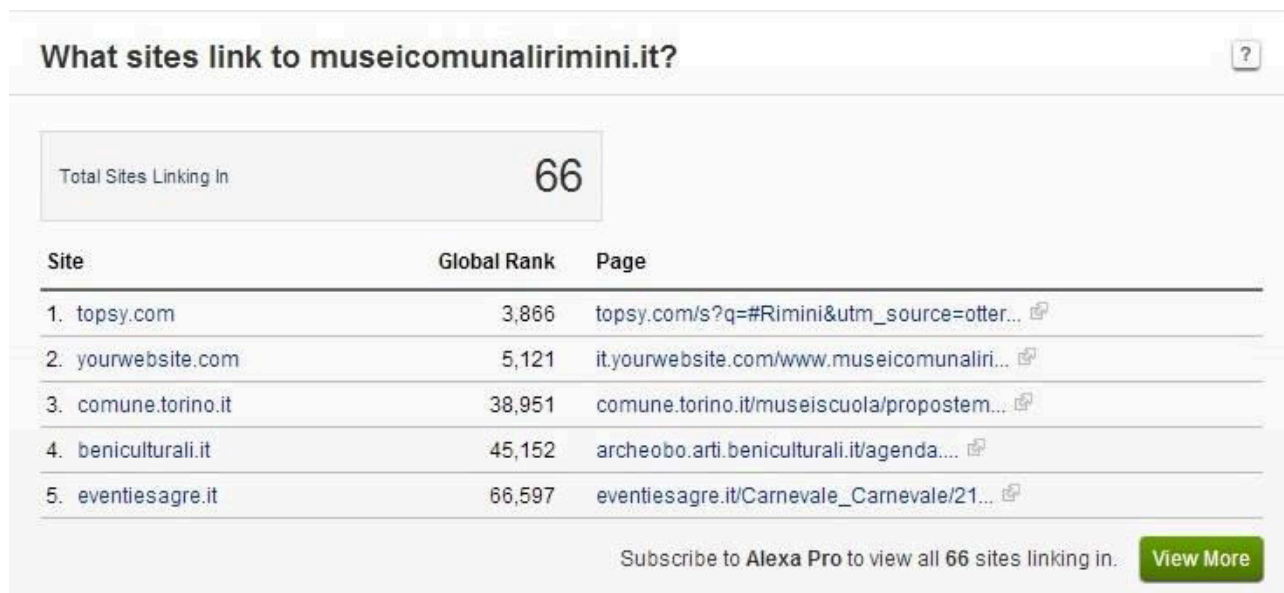
Si è già spiegato del funzionamento dello scambio di link tra più siti. In questo campo il sito dei musei della città di Rimini sembra essere abbastanza valido, come rilevato dall'analisi SEO. Sono stati analizzati anzitutto i link esterni, ovvero quelli che puntano ad altri domini. Fare riferimenti ad altri siti Web è una buona scelta e può aiutare a creare una propria autorevolezza nel settore, ma bisogna anche prestare attenzione al non averne troppi, poiché si rischia di perdere il proprio link juice¹⁴⁹. Per limitarne la perdita, è utile impiegare i file robots.txt no follow, di cui si è già ampiamente discusso. A meno che non si tratti di una resource page, avere più di 20-30 collegamenti esterni rischia di fare apparire il sito web come una link farm. Ad ogni modo, questo rischio non sembra interessare il sito oggetto di analisi, in quanto i link esterni rilevati sono 7. Secondo WooRank essi sono collegamenti a pagine di immagini o al sito del Comune di Rimini. Siti che Google identifica come relazionati a quello dei musei comunali:

- <http://www.comune.rimini.it/>
- <http://www.museoaviazione.com/>
- <http://www.teatroermetenovelli.it/>
- <http://www.sagramusicalemalatestiana.it/>

Il numero di backlink ricevuti dal sito dei musei è di 1780, secondo la funzione backlink counter di WooRank, il che è un buon risultato, equivale alla ricezione di altrettante lettere di raccomandazione per il proprio sito. L'analisi fornita da Alexa invece suggerisce che il numero di siti che si collegano direttamente al museo è di 66. Tra le varie si riportano:

¹⁴⁹ Link juice rappresenta la quantità di link presente in un sito; sostanzialmente questa quantità può essere concentrata o diluita, in base a come sono strutturati i propri link. Più link sono presenti, più i contenuti del sito sono annacquati. Altro fattore che influenza il juice è la posizione dei link (più sono vicini alla headline più hanno valore)

- www.archeobo.arti.beniculturali.it, ovvero il website della Soprintendenza per i Beni Archeologici dell'Emilia-Romagna;
- <http://www.comune.torino.it/museiscuola/propostemusei/ita/emilia/musei-comunali-di-rimini.shtml>, ovvero un collegamento dal comune di Torino riguardante la proposta di gite a musei di valore in Italia;



Per quanto riguarda invece la struttura dei link interni, essa può decisamente influenzare il ranking del sito. La costruzione di un website si basa sulla suddivisione dei contenuti in costruzioni gerarchiche di cartelle e sottocartelle; l'importanza di una pagina di un sito è determinata da quante altre pagine si riferiscano ad essa internamente (un po' come nel PageRank), segnalando quindi ai motori di ricerca quale sia la pagina più importante a parere degli sviluppatori. Gli esperti SEO consigliano di avere non più di trenta link che puntano ad altre pagine all'interno del proprio dominio. I link interni del sito dei musei di Rimini sono ben cinquantasette. Il numero deve essere chiaramente ridotto, sia perché attualmente diluisce il proprio PageRank interno, sia perché tutti questi collegamenti non possono fare altro che confondere il visitatore. Un riassunto di queste strutture di external e internal linking viene fornito da WooRank, sotto forma di grafico, che si riporta qui di seguito. Nel grafico sono visibili percentuali ben chiare:

- External links: NoFollow (0%) in quanto non sono presenti file Robots.txt
- External links: Passing Juice (21.6%) sono i 7 collegamenti verso altri siti
- Internal links (78.4%)

Gli internal links vanno necessariamente ridotti, mentre gli external links devono essere migliorati.

SEO Links

✓ In-Page Links

○○○ ○○○

We found a total of 88 link(s) including 0 link(s) to files



Anchor	Type	Juice
Vai ai contenuti	Internal Links	Passing Juice
« Home Page del Comune	External Links	Passing Juice
Image	External Links	Passing Juice
english	Internal Links	Passing Juice
MUSEI	Internal Links	Passing Juice
COME VISITARE I MUSEI	Internal Links	Passing Juice

Show more

3.2.2. Inserimento in directory

Tutti gli strumenti di analisi SEO impiegati hanno rilevato il mancato inserimento del sito Web all'interno di una directory, come DMOZ o Yahoo!. L'inserimento viene raccomandato in quanto le directory sono dei gradnsi "raccoltori", multilingua da cui i motori di ricerca possono trarre le descrizioni delle pagine Web. Tutto questo conta ai fini di ranking, per questo motivo per inserire il sito su queste directory si segnalano questi due link:

- <http://www.dmoz.org/help/update.html/> per l'inserimento nella directory di DMOZ
- <http://dir.yahoo.com/> per l'inserimento nella directory di Yahoo!

3.2.3. Inbound marketing: article marketing e comunicati stampa

Dal menu presente nella homepage del sito dei musei di Rimini non sembrerebbe possibile risalire ad alcuna forma di raccolta di articoli. Questa carenza è stata sottolineata anche da tutti gli strumenti di Analisi SEO che si sono utilizzati, sostanzialmente manca una forma di raccolta congruente di articoli scritti da un responsabile e soprattutto manca un blog personalizzato. Dalla homepage è possibile, tramite la voce "Notizie", risalire ad una raccolta di novità, ma non si va oltre questo. Ad esempio l'ultima news inserita ha titolo "Mese delle famiglie 2013", se si segue il text link si viene ri-direzionati al blog Wordpress del Centro per le famiglie della città. L'idea di poter

creare un semplice blog dove vengono raccolti tutti gli articoli relativi al museo è consigliata e a costo zero. Per quanto riguarda i comunicati stampa, certamente il museo ne fa ampio uso e vengono raccolti nel quotidiano online del comune di Rimini “Arengo”. Tramite il menu è possibile selezionare l’elenco di comunicati e visualizzare quelli relativi ai Musei Comunali. Oltre all’Arengo è possibile visualizzare tutti i comunicati stampa relativi al Festival del Mondo Antico all’indirizzo <http://antico.comune.rimini.it/>, pagina interamente dedicata all’organizzazione del rinomato festival e alla cui voce “ufficio stampa” si può selezionare “comunicati stampa” e, scegliendo il comunicato di interesse, arrivare a questo risultato:

Sei in: [UFFICIO STAMPA](#) » [Comunicati stampa](#)
21/06/2013

ANTICO/PRESENTE-FESTIVAL DEL MONDO ANTICO XV 2013 - SABATO 22 GIUGNO
Rimini - 21, 22, 23 giugno. STRADA FACENDO 2200 ANNI DI VIA EMILIA

rossegue, nella giornata di sabato 22 giugno il Festival del Mondo Antico nell’edizione intitolata STRADA FACENDO 2200 ANNI DI VIA EMILIA.

Sabato mattina alle ore 9,30 presso la Sala del Giudizio del Museo della Città Valentina Galloni (IBC), Renata Curina (Soprintendenza per i Beni archeologici dell’Emilia Romagna), Lorenza Bonifazi (Liceo “G. Cesare-M.Valgimigli”) Angela Fontemaggi e Orietta Piolanti con la collaborazione di Marzia Ceccaglia (Musei Comunali Rimini) presentano le attività degli studenti coinvolti nel progetto del Liceo “G. Cesare-M.Valgimigli” e dei Musei Comunali di Rimini, fra i vincitori della II Edizione di “Io amo IBC” con il progetto Le pietre raccontano.

A seguire gli studenti daranno voce alle pietre del Lapidario romano guidando il pubblico fra le iscrizioni esposte nel giardino del Museo.

Alle ore 10, presso la sede della Banca Malatestiana in Palazzo Ghetti (via XX Settembre 63), verrà inaugurata l’esposizione archeologica Dalle origini del Borgo a Palazzo Ghetti, 25 secoli di storia, a cura della Soprintendenza per i Beni archeologici dell’Emilia Romagna. Intervengono Filippo Maria Gambari (soprintendente), Enrica Cavalli (presidente di Banca Malatestiana), Alessandro Mori (Cumò Mori Roversi architetti). Presentazione del percorso espositivo a cura di Maria Grazia Maioli e ad Arte di Luca Mandolesi. Inaugurazione e apertura straordinaria sabato 22 giugno 2013, 10-13 e 15-18. Da lunedì 24 giugno la sala espositiva sarà aperta solo su appuntamento: Tel. 0541 315811

Per i riminesi era semplicemente “il Palazzo dei fiammiferi”. In realtà l’edificio che dal 1857 al 1908 ha ospitato la “Fabbrica di zolfanelli fosforici” del Cav. Nicola Ghetti ha sempre avuto un profondo legame con la storia di Rimini.

Quanto profondo e con quanta storia, lo si è appreso però solo grazie agli scavi archeologici promossi tra il 2009 e il 2010 dalla Banca Malatestiana, proprietaria dell’immobile, sotto la direzione scientifica della Soprintendenza per i Beni Archeologici dell’Emilia-Romagna. Le indagini archeologiche, effettuate nelle due aree cortilizie, hanno prima intercettato le strutture legate alla fabbrica ottocentesca, e poi portato in luce strutture medioevali del Borgo San Genesio (oggi San Giovanni): case, botteghe, magazzini e cortili con materiali che hanno aperto una finestra sulla vita quotidiana della Rimini malatestiana tra il XIV e XV secolo.

Il Borgo era stato costruito sopra una necropoli romana del III-IV sec. d.C., lungo la via Flaminia (oggi via XX Settembre). Sotto gli strati sono infine emersi i fossi di bonifica del territorio, realizzati dai Romani a partire dalla metà del III secolo a.C.: questi fossi (riempiti di terreno) hanno restituito vasellame e altri reperti importanti per la storia di Rimini.

L’allestimento di uno spazio espositivo permanente nel Palazzo Ghetti, voluto da Banca Malatestiana e accolto dalla Soprintendenza è un ulteriore passo all’interno di un percorso di tutela e conoscenza, iniziato con lo scavo archeologico.

Tornando al Museo si potrà seguire Emilia una strada per il Nord. Giovanni Brizzi (Università di Bologna), Filippo Maria Gambari (soprintendente Beni archeologici Emilia Romagna) e Iacopo Ortalli (Università di Ferrara) analizzano la via Emilia come paradigma di una strada consolare su cui muove la romanizzazione. Se ne osservano i risvolti di via militare, di asse su cui cresce un importante paesaggio monumentale: di realtà ancora fondante per la quale l’indagine archeologica e gli allestimenti museali si propongono come percorsi di conservazione e di comunicazione del patrimonio.

3.2.4. Utilizzo del SMO

Per quanto riguarda i social media, WooRank riassume con una figura la presenza dei musei di Rimini sui network:



L’impatto cosiddetto “social”, è essenziale in certi settori, e quello museologico non è da meno se si

osserva tutto lo spazio dedicato a questo settore da grandi nomi quali il British Museum o il Louvre, ma anche le varie sedi Guggenheim in tutto il mondo. Bisogna accertarsi di essere in grado di raggiungere i propri utenti su diverse piattaforme, a cominciare da Facebook, ma anche Twitter, LinkedIn, Google+ e programmi di social bookmark quali StumbleUpon o DiggIt. Bisogna prestare molta attenzione e sforzi nel creare un legame con gli utenti social e creare una fan base consistente. Tornando al sito dei musei di Rimini, sembrerebbe che l'unico social network in cui si è attivi sia Facebook:



Come si può vedere la pagina è abbastanza seguita con 547 gradimenti e 25 citazioni online, sempre sul social network. È anche importante segnalare che la pagina viene aggiornata quasi quotidianamente e quindi il personale museale è piuttosto attento a diffondere con foto, video e notifiche di eventi, i principali avvenimenti di ambito artistico della provincia.

Sfortunatamente, questo è l'unico network impiegato, non figurano altre pagine ufficiali dei Musei Comunali di Rimini. Su Twitter si può vedere che spesso il museo e le sue collezioni vengono citate, e questo rientra sempre nell'ambito dei backlink di cui si è già parlato. Si è deciso di riportare in figura quanto visualizzato se un utente effettua la ricerca "musei Rimini" su Twitter.



Per quanto riguarda il social bookmarking, si sono trovate le icone di digg.com e delicious.com sulla homepage del sito, come a rimandare l'utente ai profili del museo contenuti in quelle pagine, ma in seguito ad ispezione i collegamenti si sono rivelati nulli. È quindi anche consigliato di registrare i Musei Comunali o di aggiungere la homepage del sito ufficiale ad almeno queste applicazioni:

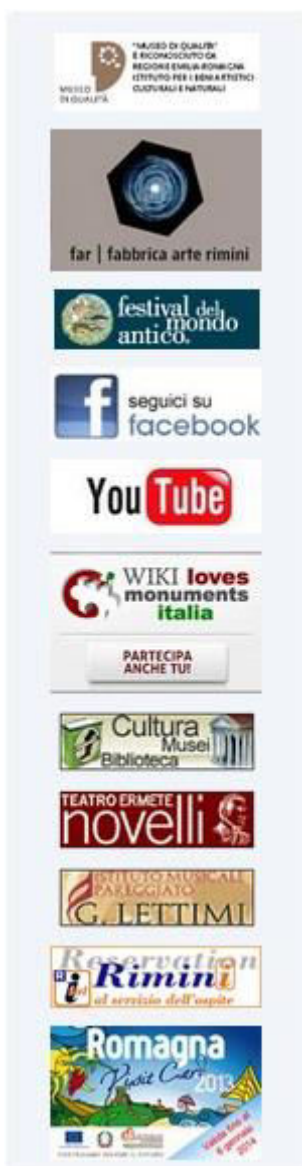
- Twitter, servizio di microbloggin, consentirebbe allo staff dei musei di tenersi sempre in contatto con chi si segna come “follower” della pagina. I posti condivisi possono avere una lunghezza massima di 140 caratteri ed è possibile per gli utenti “retwittare” le notizie che a loro paiono più importanti, consentendo una maggiore risonanza dei propri post;
- Google+, social network che permetta la condivisione di foto, chat, messaggi e contatti grazie alle cosiddette “Cerchie”. Più cerchie si riescono ad ottenere più i propri messaggi avranno diffusione; oltre all'aspetto social si ricorda l'avvento di Google Plus Your World, per cui sono stati integrati i cosiddetti “+1” nei risultati di ricerca e quindi scegliere di implementare un profilo Google Plus è estremamente rilevante ai fini di ranking.

Venendo ora al discorso social bookmarking, si tratta di siti sviluppati per fare in modo che gli

utenti possano salvare i proprio siti preferiti e renderli disponibili da qualsiasi punto di accesso, indipendentemente dal computer usato. Sulle pagine del sito dei musei comunali si sono trovate icone di digg.com (sito di social news che consente allo staff museale di condividere i propri contenuti con tutti) e delicious.com ma i link non sono attivi ed effettuando una ricerca sui due social bookmark non si trovano pagine relative ai musei riminesi. Per questo motivo si consiglia di provvedere all'iscrizione su queste due pagine, e anche su StumbeUpon o Diigo.

3.2.5. Banner advertising e e-mail marketing

Tutti i siti collegati e gestiti dal comune della città di Rimini, presentano nella schermata dalla homepage, nella colonna di sinistra, generalmente sotto al menu principale una lista di banner diversi, di rimando ai vari siti. Per esempio, sulla homepage dei musei comunali si trovano questi banner:



Questi banner consentono di collegarsi ai siti correlati a quello dei Musei Comunali, come per esempio

- la pagina della Soprintendenza;
- la Fabbrica Arte Rimini;
- il Festival del Mondo Antico;
- la pagina dei musei su Facebook;
- il canale YouTube dei Musei;
- la pagina di Wikipedia dedicata ai monumenti di Rimini;
- la pagina del Comune di Rimini servizi/città/cultura;
- il Teatro Ermete Novelli;
- l'Istituto Musicale G. Lettimi;
- la pagina di riminireservation.it;
- la pagina dedicata alla Romagna Visit Card, convenzione per visitare tutti i principali monumenti romagnoli a prezzi vantaggiosi;

Per quanto riguarda l'e-mail marketing, i Musei Comunali di Rimini non presentano un modulo di compilazione per iscriversi ad una newsletter. Si presuppone che azioni di e-mail marketing vengano effettuate direttamente dallo staff su liste di e-mail pre-acquisite, ma senza dare la possibilità di iscriversi agli aggiornamenti (al di là degli RSS feed) è difficile poter allargare il proprio giro di conoscenze.

Giunti al termine di questo capitolo si pensa di aver mostrato adeguatamente quali siano le principali attività di ottimizzazione che occorre implementare sia a livello interno, sulla struttura del sito museale, sia invece a livello esterno, di popolarità e visibilità sul Web. L'utilizzo dei vari software di SEO Analytics si è rivelato molto prezioso ai fini di ottenere una diagnosi il più vicina possibile alla situazione effettiva. La loro importanza risiede nel fatto che, oltre a fornire un'analisi generica della struttura e del valore SEO del sito, questi software sono anche in grado di assegnare un punteggio valutativo. Si è scelto di riportare in particolare in figura il punteggio assegnato da WooRank, in quanto rappresentante il giudizio in gran parte condiviso dagli altri software.

La votazione assegnata al sito del Museo della Città è di 57.5%. Si tratta di un punteggio superiore alla media e quindi ampiamente sufficiente, ma lontano dall'eccellenza che si auspica si possa un giorno raggiungere, magari anche grazie al piccolo aiuto fornito dalle osservazioni contenute in questo elaborato. Oltre al punteggio e al rendering della homepage è visibile una tabella riassuntiva dell'analisi svolta; in questo caso si legge che il sito risulta ottimizzato per ventidue fattori esaminati, potrebbe effettuare miglioramenti su quattro elementi, mentre devono essere necessariamente riviste almeno tredici impostazioni.



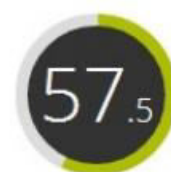
Review of [www.museicomunalirimini...](http://www.museicomunalirimini.it)

Generated on September 25 2013 2:35 PM

Successfully passed 22

Room for improvement 4

Errors to fix 13



Get short URL

Si è già detto in apertura dell'obiettivo di questo elaborato, ovvero la dimostrazione nella modalità più oggettiva possibile del perché, al giorno d'oggi, sia assolutamente necessario effettuare investimenti nelle tecniche SEO. Esse sono fondamentali per portare ad un drastico aumento di visite al sito e quindi questa visibilità ottimizzata può solo avere conseguenze positive per i musei,

ai fini della fornitura di un servizio adeguato alla cittadinanza. Tramite l'implementazione di alcuni suggerimenti dell'ultimo paragrafo si può aumentare il contatto diretto tra i cittadini e i musei, in particolare fornendo risorse al visitatore. Si pensi alla possibilità di inserire un banner per l'iscrizione alla newsletter al conseguente ottenimento di contatti e-mail, la partecipazioni a eventi sponsorizzati e magari anche alla realizzazione di un negozio online dove è possibile acquistare libri e altri oggetti con il marchio dei Musei della Città. Fornire qualcosa di unico all'utente in cambio di una qualche azione (che sia di registrazione o un aiuto alla diffusione di un evento) migliora sia l'aspetto interattivo che quello della fidelizzazione. Quest'ultimo è essenziale, così come la necessità di inserirsi in un aspetto di quotidianità agli occhi degli utenti, che devono arrivare a percepire il museo come un'entità in continua evoluzione e pronta a fare quel passo in più per rimanere in contatto con gusti e esigenze di tipologie di utenza diversificate. Questo aspetto potrebbe portare anche alla scoperta di nuovi target o settori di business precedentemente non considerati, il tutto legato anche alla capacità di effettuare i giusti investimenti e le giuste scelte in campagne di AdWords e di analisi delle parole chiave.

Se a questo discorso si aggiungono riflessioni relative alla direzione di sviluppo delle tecniche SEO, che pare andranno verso un'ottica sempre più local e geo-localizzata (si pensi a Google Glass) si capisce la necessità di un approfondimento di questa tipologia di analisi. In particolare l'aspetto social diventerà sempre più integrato al web searching, unendo indissolubilmente media diversi come foto, video, recensioni ecc. e rispondendo ad esigenze ancora più personali. Soprattutto l'aspetto più approfondito dai motori di ricerca riguarda la loro capacità di effettuare ricerche su basi semantiche piuttosto che analizzando staticamente keyword separate. Questo sarebbe utile anche ai fini di posizionamento a livello internazionale, basti pensare al fatto che attualmente il ranking si deve ottenere in nazioni con parole chiave differenti e in mercati raramente equivalenti. L'implementazione dell'aspetto semantico sarebbe un passo monumentale verso quell'obiettivo di applicazione dell'intelligenza artificiale ai motori di ricerca, ma per poterne fare pieno uso e trarne il massimo risultato bisogna fornire al proprio sito contenuti validi e ottimizzati. Si presenta un esempio pratico, se oggi si dovesse cercare "Musei Comunali di Rimini" su Google, il motore ci fornirebbe il sito ufficiale e qualche altro URL di collegamento alla pagina; l'obiettivo ultimo di questi sviluppi tecnici sarebbe che in futuro un utente potrebbe digitare "quali sono le dieci cose essenziali da visitare in Emilia-Romagna?" e in automatico il search engine fornirebbe nella sua pagina, un elenco (possibilmente includendo il sito riminese) senza presentare una lista di collegamenti alle varie pagine. Si capisce che la fornitura di contenuti di alto valore diventa essenziale e per fare ciò bisogna massimizzare sia l'aspetto on-page, che quello off-page,

aumentando il circolo di collegamenti alla pagina, onde mostrare a Google la sua rilevanza e autorità. Questo obiettivo può essere raggiunto un passo alla volta, magari cominciando proprio dall'applicazione dei suggerimenti raccolti in questo elaborato, piccole migliorie che sommate e unite ad una ricchezza di contenuti di buona qualità, sicuramente migliorerebbero le performance del sito, anche in preparazione all'avvento delle nuove tecnologie.

Nel prossimo capitolo si intende fare una ricapitolazione delle principali modifiche e implementazioni da apportare, seguite da un modello in grado di illustrare esattamente, a livello monetario, quanto possa essere conveniente avere un sito ottimizzato dal punto di vista SEO.

LA VALUTAZIONE DELL'INVESTIMENTO

In questo capitolo si intende avvicinare l'argomento oggetto di questo elaborato da un punto di vista più pratico. Quello che si vorrebbe fare è anzitutto ricapitolare quelle che sono le azioni più urgenti da intraprendere per l'ottimizzazione del sito, fornendo anche le tempistiche di svolgimento dei vari compiti, e successivamente dimostrare mediante l'utilizzo di alcuni grafici e modelli il legame tra l'applicazione di alcune tecniche di ottimizzazione SEO al website e il ritorno economico che deriverebbe direttamente ai Musei Comunali. Purtroppo, non essendo in possesso di alcuni valori reali, quali la capacità di investimento di cui si potrebbe disporre, o il valore economico di una visita al sito web, ci si trova costretti a dover effettuare delle stime e a proporre quelli che sono i risultati più probabili.

1. Le azioni da effettuare con più urgenza

Si è scelto di riportare le azioni di ottimizzazione tramite il supporto di una "gant chart". Questa tipologia di tabella è stata sviluppata ad inizio Novecento da Henry Gantt ed è uno strumento molto utile per illustrare sinteticamente, ma correttamente, quelle che sono le tempistiche di realizzazione di un progetto di qualsiasi tipo. Nel caso che qui si tratta, si illustrano tutte le attività divise tra on-page e off-page. Le prime riguardano le azioni da intraprendere sulla struttura della pagina, a livello di HTML e porzioni di codice, mentre le seconde riguardano più la popularity del sito e i collegamenti al resto del web. La gantt chart consente di illustrare una stima delle date d'inizio e fine di un progetto, dividendole tra le varie azioni e mostrando anche i legami di sequenza tra alcune di queste¹⁵⁰.

Tabella gantt

Le stime temporali sono state calcolate ipotizzando liberamente che il personale impiegato per svolgere queste azioni sia di una, massimo due, persone. Queste persone è difficile che possano dedicare il loro intero orario lavorativo a questo tipo di studio, per cui si è cercato di fare una stima temporale tenendo conto che lo staff vi possa dedicare massimo un'ora o due al giorno. Come è facilmente visibile dalla tabella gantt, esistono interventi la cui messa in pratica richiede al massimo un mese e altri che devono essere portati avanti in via continuativa per tutto l'anno.

Iniziando dalle tecniche on-page, si è pensato di inserire nel grafico come prioritari diversi interventi. L'effettuazione di una campagna AdWord è di importanza tale per cui non può essere iniziata e poi sospesa dopo un breve periodo, è necessario curarla continuativamente se si vuole capire l'efficacia delle scelte realizzate per determinate parole. Al contrario, l'acquisizione dei

150 http://www.mindtools.com/pages/article/newPPM_03.htm

diversi domini può essere risolta in un mese, così come l'implementazione dei robots meta tag e della sitemap in formato XML. La correzione dei meta tag e l'inserimento di microformati possono richiedere un po' più di tempo, soprattutto se il personale non è preparato a livello di codici HTML e altri linguaggi specifici, ma si tratta di interventi che devono essere subito intrapresi, in quanto il loro impatto sul ranking è di indiscutibile rilevanza. Successivamente, si è pensato di inserire le modifiche da effettuare relative all'ottimizzazione della pagina, quali riduzione di peso, tempi di caricamento e implementazione di cached pages. La riduzione dei tempi di caricamento della pagina si pensa che possa prendere più tempo rispetto alle altre azioni, e può essere accompagnata dalle migliorie nella gestione delle immagini e nell'inserire più text content per rendere il sito più coinvolgente e interattivo rispetto all'eccessiva schematicità attuale. Naturalmente la creazione di contenuti rilevanti si è limitata a tre mesi come implementazione, ma poi dovrebbe essere continuata nel tempo, ogni volta che si effettuano delle modifiche. Per ultime si sono inserite la correzione delle URL e la creazione di una versione mobile, queste azioni non hanno la massima priorità, soprattutto se paragonate alle altre, però richiedono tempo e capacità tecniche non indifferenti, soprattutto la seconda delle due.

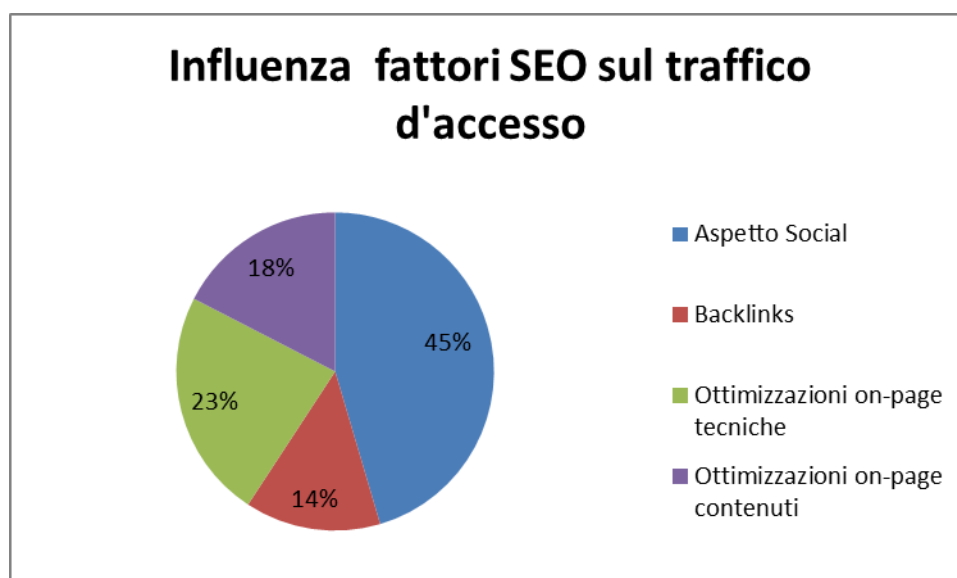
Dal punto di vista off-page, le azioni che si sono inserite per prime sono quelle di attivazione dei profili Twitter, Google+ e social bookmark vari. Chiaramente, ai fini della creazione di un buon profilo ufficiale, arricchito di attributi e immagini personalizzate, può essere necessario massimo un mese (probabilmente anche meno) di lavoro, ma la corretta manutenzione e aggiornamento dei contenuti deve continuare per tutta la vita dei Musei Comunali; non si può rischiare di perdere utenti o venire meno alle aspettative del pubblico. Successivamente si è pensato di inserire la creazione di un blog che dovrebbe raccogliere anche gli articoli relativi ai musei e creare più collegamenti al sito e condivisioni della pagina, la sua importanza è essenziale. Allo stesso livello di priorità si è pensato di inserire anche tutto quell'insieme di azioni che riguardano il collegamento pratico tra il sito web museale e la struttura vera e propria; di queste si parlerà nello specifico brevemente. Successivamente si consiglia la creazione di un banner utile alla registrazione di nuovi utenti al servizio di newsletter, questa implementazione porta anche ad un allargamento della propria mailing list. Azione che deve essere prolungata nel tempo è quella di article marketing, mentre per l'inserimento in directory si è segnato una tempistica massima di un mese, anche se in realtà probabilmente ci si potrebbe impiegare di meno. Infine si è inserita la diminuzione dei collegamenti interni e l'implementazione di outbound linking, il cui espletamento richiede tempo per la ricerca dei migliori siti a cui collegarsi.

2. Influenza dei fattori SEO sul traffico d'accesso

Secondariamente si è pensato di illustrare quella che è l'influenza diretta dei fattori SEO sull'aumento del traffico sul sito Web. A questo proposito si è ripreso in mano lo studio sui fattori di posizionamento di Google illustrato al termine del primo capitolo e si è creato un grafico. Tramite esso si desidera esprimere esattamente l'influenza delle azioni di ottimizzazione sull'ottenimento di visualizzazioni del website e quindi conseguenti visite aggiuntive. Le singole azioni elencate nella gantt chart sono state riassunte in macro-categorie, riprendendo la modalità effettuata dallo studio, per cui si avrà:

1. Aspetto social, con l'implementazione di tutti i profili Google+, Twitter, ecc.;
2. Backlinks, il loro numero, la percentuale di stop words, il rel="nofollow" e la loro provenienza;
3. Ottimizzazione on-page dal punto di vista tecnico, quindi i meta tag e le modifiche al codice HTML
4. Ottimizzazione on-page dal punto di vista dei contenuti, quindi le parole chiave, i collegamenti esterni, ecc.

Sostanzialmente da questo grafico è deducibile quanta attrazione può catalizzare ogni categoria di Search Engine Optimization.



Ora che si sono illustrate le principali azioni di ottimizzazione che è necessario implementare sul sito web dei musei riminesi, si pone un interrogativo importante, volto a capire il legame tra sito web e visitatore fisico. Infatti, nel caso di un sito Internet di un museo pubblico, non è possibile

ottenere dei ritorni economici per ogni accesso effettuato da un utente. Il modo più comune tramite cui un sito web riesce ad ottenere ritorni economici, si è già detto, è inserendo delle pubblicità direttamente on-page che, in seguito ai clic effettuati da utenti, garantiscono un guadagno. L'esempio più volte fornito è quello di Google AdSense, che è anche il più diffuso. Tuttavia, per i Musei Comunali di Rimini, il vero ritorno economico si ha quando il visitatore fisicamente si presenta e compra il biglietto alle casse, non esistono pubblicità sui siti comunali. Allora per quale motivo bisognerebbe investire nell'ottimizzazione del sito se, almeno apparentemente, non esiste un collegamento diretto tra visitatore e sito web? Ovviamente questo collegamento esiste, eccome.

3. Il legame tra user online e visitatore museale

Nel momento in cui un utente effettua l'accesso ad un sito museale si aspetta di trovare determinate informazioni, immediatamente e senza difficoltà:

1. un percorso chiaro di massimo 2-3 clic dalla homepage per raggiungere dettagli logistici quali localizzazione, orari d'apertura e costi d'accesso
2. informative chiare riguardo eventi e mostre, sempre facilmente raggiungibili, evitando agli utenti di rimanere bloccati su grovigli di link interni;
3. servizi di prenotazione online;
4. rappresentazione della collezione museale online, con una buona qualità d'immagine;

Queste sembrano informazioni e dettagli scontati, soprattutto alla luce di nuove tecnologie e nuove possibilità di presentazione di un sito, però è sconcertante quanto numerosi siti museali non contengano percorsi chiari per raggiungerle, e la difficoltà di navigazione sembra quasi presagire quella che sarà poi la difficoltà di visita o di identificazione di un percorso di visita all'interno delle sale. Tutto il resto è un surplus, ma queste info devono essere subito reperibili. A questo proposito il sito dei Musei della Città di Rimini, presenta queste informazioni, anche se sono necessari svariati clic per raggiungere le pagine, divise in maniera estremamente schematica. Per quanto riguarda invece servizi aggiuntivi, sul sito non sono presenti molti contenuti. Tra i più importanti si segnala il collegamento al catalogo della collezione, che è interamente affidato a quello dell'IBC di Bologna. Sembra che i reperti archeologici non siano ancora stati inseriti, mentre i dipinti e le tele della pinacoteca risultano tutti catalogati online. Ma perché sono importanti questi servizi aggiuntivi? La loro rilevanza sta nella capacità di attirare l'utente prima e dopo della visita museale. Un conto è che un turista in procinto di visitare la città di Rimini scelga di interessarsi alla proposta culturale della città e quindi cerchi e visualizzi il website dei musei; ma un altro è se in seguito alla visita l'utente scelga comunque di rimanere in contatto con la struttura, magari registrandosi alla newsletter o inserendo il sito tra i preferiti. L'obiettivo più importante da conseguire da parte dello

staff, a parere di chi scrive, dovrebbe essere quello di spingere gli utenti a visitare il sito in maniera continuativa, anche dopo l'averne visualizzato la collezione. Ciò sarebbe interessante, e utile, perché consentirebbe agli utenti di rimanere aggiornati sulle novità e di usufruire di contenuti unici e visibili solo online. Ad esempio, nel blog del museo Tate Modern di Londra, alla domanda posta agli utenti del sito da un membro dello staff: "cosa desiderate dal sito di un museo?" alcuni utenti hanno espresso curiosità per le scene "dietro le quinte" delle varie installazioni¹⁵¹. Ovvero desidererebbero che venissero mostrati i momenti antecedenti all'inaugurazione e poi successivi alla chiusura, finanche ai lavori di "preservation" delle opere. Insomma, informazioni uniche e impossibili da reperire alternativamente. Questo andrebbe chiaramente oltre le normali funzioni di un sito web, è quasi se i nuovi siti ottimizzati di alcuni musei diventassero delle destinazioni essi stessi, con tutti i contenuti esclusivi presenti e i supporti alla ricerca. Nel 2000 era necessario essere presenti sul web per fare in modo di essere rintracciabili tramite Google, ma adesso il web sta cambiando in un ambiente informativo socialmente contestualizzato, ed è importante quindi essere rintracciabili, mediante i social network dei singoli utenti. Attualmente si effettua l'accesso a Internet per ottenere delle informazioni ed è questo il motivo che ha spinto i musei a fornirsi di siti web. La prima motivazione è quella di fornire informazioni logistiche basilari di cui si è già scritto, sono quelle informazioni che si raccolgono durante la preparazione di un viaggio, la seconda riguarda fornire qualcosa in più ai proprio utenti. I dipartimenti di marketing hanno realizzato che una grande percentuale di persone utilizza i search engine per trovare cose interessanti da fare nel tempo libero e questo è un settore in cui le istituzioni museali devono riuscire gradualmente ad inserirsi.

Quanto scritto finora costituisce l'aspetto più culturale e educativo del legame che caratterizza la realtà museale e il visitatore del website, ma è chiaro che essere in possesso di un buon website non possa fare altro che aumentare la popolarità e l'interesse per la struttura vera e propria. La modalità migliore e diretta per evidenziare questa relazione sarebbe quella di fornire all'utente che ha visitato il museo un questionario, chiedendo in che modo sia venuto a conoscenza della struttura e in particolare se ha visitato il sito internet per:

- visualizzare orari e altre informazioni logistiche;
- vedere le mostre attualmente presenti;
- visualizzare determinati pezzi di collezione a cui si è interessati;

Si potrebbe effettuare un sondaggio contenente queste richieste direttamente all'uscita del museo, o sulla pagina Facebook ufficiale dei Musei della Città di Rimini, o sul website. Sulla homepage si

151 <http://www.tate.org.uk/context-comment/blogs/tate-debate-what-do-you-want-museum-website>

potrebbe creare un banner interattivo tramite cui l'utente può segnalare se, in seguito alla visualizzazione del sito, sarebbe interessato a visitare il museo o se ha già effettuato visite precedenti e, in questo caso, per quale motivo stia visualizzando nuovamente la pagina internet. Le risposte a questi quesiti sono essenziali per poter coltivare sempre più il legame con i visitatori, prima durante e dopo la visita. Un'altra interessante modalità per approfondire il legame tra sito e museo potrebbe essere quella di offrire dei coupon di buono sconto o sul biglietto d'ingresso o, ancora meglio, sul merchandise venduto all'interno del bookshop. Se poi si pensa ad un potenziale negozio online, le possibilità sono ancora maggiori.

4. L'investimento nelle tecniche SEO

Si intende ora illustrare un particolare modello, molto utile per capire per quale motivo l'istituzione museale dovrebbe scegliere di effettuare un investimento nello sviluppo di tecniche SEO da applicare al sito museale. Questo può essere realizzato o mediante l'assunzione di nuovo personale che abbia una formazione e interesse in questo campo o nell'istruire adeguatamente, mediante corsi di formazione, lo staff attuale in modo che possa interessarsi e agire anche in questo settore. Il modello analizza determinate parole chiave, valutandone l'attuale volume di ricerca mensile su Google e, in base al ranking che la pagina ottiene per ognuna di esse, determina il traffico di accessi al sito dei musei di Rimini. Dopodiché è sufficiente fornire un obiettivo di ranking che è valutato su basi analitiche quali backlink esatti, competizioni e valori effettivi di ricerca (di cui purtroppo attualmente non si dispone nel dettaglio, in quanto informazioni riservate al webmaster del sito museale) e il modello calcola automaticamente l'aumento del traffico in relazione al nuovo ranking, sempre diviso per parola chiave. Per quanto riguarda la struttura del file, nella prima colonna si trova un elenco di parole chiave, alcune sono quelle già presentate nel capitolo precedente, tra keyword density e keyword meta tag, ma la maggior parte si sono aggiunte mediante l'utilizzo dell'utilissimo tool Ubersuggest.com, un valido suggeritore di parole chiave. Venendo alle colonne numeriche:

- Volume (Exact) fa riferimento al traffico di ricerche mensili fornito da Google per la parola chiave in questione; sostanzialmente è la domanda di mercato per quella determinata keyword. I dati sono stati reperiti da Google AdWords, Google Keyword Tool e Google Keyword Planner;
- Current Tot Traffic si riferisce al numero di visite, sempre mensile, che il sito ha ottenuto in seguito ad un determinato ranking per un certo termine di ricerca;
- Current Ranking contiene il posizionamento nella SERP che il sito ha ottenuto per una certa keyword digitata;

- Target Ranking è il posizionamento obiettivo che si intende conseguire in seguito all'ottimizzazione on-page e off-page del sito dei Musei Comunali di Rimini. È stato scelto un valore simbolico di 3, perché è impossibile stabilire quale sia esattamente il target che si potrebbe raggiungere senza disporre dei dati analitici esatti necessari;
- Est. CTR è la click-through rate stimata, ovvero una percentuale indicante il successo di una campagna pubblicitaria per un sito web data dal rapporto in percentuale tra i click che il banner pubblicitario ha ottenuto e il numero di volte in cui esso è stato mostrato, detto impressions;
- Est. New CTR è la nuova click-through rate stimata per il nuovo ranking;
- CTR ratio è il rapporto tra le due percentuali, indica il miglioramento nel numero dei clic;
- Est. New Traffic è il calcolo del nuovo traffico d'accesso al sito in seguito all'implementazione del nuovo ranking;
- Est. Traffic Increase è la differenza tra il numero ottimizzato di visite e quello precedente al miglioramento del sito, indica proprio il numero di visite in più in seguito all'applicazione di tutte le azioni migliorative riassunte nella gantt chart di inizio capitolo;

Quindi per ogni riga si calcolano tutti questi dati in riferimento alle diverse parole chiave, mentre nell'ultima sono contenuti i valori totali. In questa stima che si è fatta, si calcola che l'implementazione delle tecniche SEO tramite un investimento che si può stimare di 1800 euro mensili e l'utilizzo di campagne per tutte quelle parole chiave può portare ad un aumento di visite al sito da 1212 a circa 6613. A questo punto si propone un ragionamento, ovvero che è lecito supporre che di queste 6613 visite aggiuntive al website del sito dei musei riminesi, una piccola percentuale si possa tradurre in una effettiva presenza alle casse. Se anche si trattasse dell'1%, sarebbero comunque 60 biglietti mensili in più. Se si considera che un biglietto intero costa 6€, sarebbero 3600€ al mese in più. Questo significa che:

$$\mathbf{ROI} = 3600/1800 = 2 = 200\%$$

Certo si tratta sempre di valori ottenuti dalla stima di un miglioramento di ranking, ma comunque questo ritorno giustificerebbe un eventuale investimento in questo campo.

La scrittura di questo capitolo era intesa per dimostrare come l'applicazione delle tecniche SEO possa garantire dei risultati efficaci per il sostentamento del museo e anche efficienti, perché i costi potrebbero essere molto contenuti e limitarsi solo alla retribuzione del personale specializzato o nell'educazione dello staff attualmente presente in queste materie informatiche.

CONCLUSIONI

Nel corso di questo elaborato si ritiene di aver dimostrato effettivamente come l'analisi a livello SEO di una pagina web sia attualmente vitale per lo sviluppo e il sostentamento di un'istituzione museale. La situazione del sito dei Musei della Città di Rimini è inseribile in un livello di sufficienza, che molto ha da migliorare e studiare. In chiusura si pensa di fare una piccola ricapitolazione da un punto di vista molto pratico, di quelle che sarebbero alcune azioni da intraprendere per vedere miglioramenti sostanziali e in breve tempo. Anzitutto quello che sarebbe più utile fare sarebbe analizzare il traffico web del sito museale, utilizzando direttamente Google Analytics che dall'analisi di cui al capitolo III risulta già installato. Secondariamente è necessario fornire una preparazione adeguata su web analytics e tecniche SEO, sarebbe ideale che almeno parte del personale frequentasse corsi di preparazione su questo argomento, nell'impossibilità di fare ciò si potrebbero richiedere servizi di consulenza a aziende esterne. Attività di studio su web analytics, utilizzo dei social network e creazione dell'immagine e del brand sono essenziali per la prosperità di qualsiasi azienda, e recentemente queste tecniche sono state impiegate con grande successo da numerosi musei di rilevanza internazionale. Tramite l'utilizzo di queste azioni si può dare un valido aiuto all'autosostentamento dell'istituzione museale, se poi esse si accompagnano alla cura e allo studio di contenuti adeguati, si rivelano essenziali per garantire un numero maggiore di visite al sito, accompagnate da alti tassi di conversione in potenziali visitatori fisici e maggiore fedeltà al museo. Dopo un breve periodo in cui si sono iniziate le prime azioni di ottimizzazione, si raccomanda di svolgere una prima revisione generale a livello SEO. Questa revisione dev'essere anzitutto dedicata all'analisi della struttura e architettura del Web, rispetto a quella che è la propria SEO checklist, ovvero la propria lista di azioni importanti da svolgere. Un buon esempio di checklist potrebbe essere quella svolta nel capitolo III e riassunta nella gantt chart del capitolo IV con tanto di stime tempistiche. In seguito anche ai risultati del primo controllo bisogna poi indirizzare la propria azione di Search Engine Optimization, a livello on-page.

Lo sviluppo di nuove tecnologie di web 2.0 ha presentato ai professionisti che lavorano in ambito museale nuovi modi di intendere e nuove possibilità di connettersi con i propri pubblici online e soprattutto coinvolgere profondamente le persone nell'ambiente museale. Come già scritto precedentemente, si tratta sempre più di inserire non il visitatore nella vita del museo, ma far diventare il museo una presenza consistente nella vita del visitatore. Ciò è reso possibile grazie ai nuovi strumenti interattivi e ai nuovi livelli di connettività che consentono ai visitatori di accedere al museo prima ancora di averlo fisicamente visitato o di continuare a frequentarlo anche una volta tornati a casa. È stato sottolineato da alcuni studi che le motivazioni che possono spingere una

persona a recarsi fisicamente ad un museo e le motivazioni che possono spingere un utente ad accedere ad un sito web museale sono essenzialmente diverse: “Physical museum goers are seeking experiences-learning, experiences perhaps, but experiences non the less. In contrast the Internet was created for resource sharing and communication. This distinction shapes the current differences in motivation in the two ventures”¹⁵² la stessa ricerca sostiene che i visitatori decidono di recarsi ai musei perché queste istituzioni sono delle cosiddette “identity shapers”, ovvero creano un’identità culturale. Si approfitta in chiusura di questa validissima argomentazione per sottolineare che con questo elaborato non si vuole consigliare ai musei di perdere completamente la loro funzione pratica e spaziale, la loro essenza, che è quella di contenere e *mantenere vive* tradizioni e contenuti storici. L’organizzazione di mostre e esposizioni fisiche è essenziale come generatrice di identità, ed è fondamentale nella creazione di un senso di appartenenza culturale per il cittadino e la comunità in cui si muove l’istituzione culturale; è evidente che il web non è assolutamente in grado di fare altrettanto e di poter incarnare gli stessi valori. È vero che la proliferazione di smartphone e altri strumenti mobile sempre più sofisticati, ha incoraggiato e tuttora spinge gli addetti ai lavori a pensare a nuovi strumenti di appropriazione e diffusione di contenuti, si pensi per esempio al museo di arte moderna di San Francisco che ha realizzato una guida alla sua collezione permanente con tantissimi extra chiamata “Making sense of modern art”¹⁵³. Le tecnologie web 2.0, ovvero quelle che favoriscono e facilitano la condivisione sicura di informazioni (wiki, blog, applicazioni di condivisione di musica/video/libri, social bookmarking, RSS feed, ecc.) forniscono grandi opportunità alle istituzioni museali, per aprirsi al pubblico e mantenere vive e fruibili le proprie collezioni. Si vuole ricordare, a questo proposito, la definizione aggiornata alla ventiduesima assemblea generale dell’ICOM, tenutasi a Vienna nel 2007, del termine “museo” che è la seguente:

*“A museum is a non-profit, permanent institution in the service of society and its development, open to the public, which acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study and enjoyment”*¹⁵⁴

Il museo non è solo un luogo dove le collezioni devono essere esposte, è necessario cercare di comunicare e fornire nuovi collegamenti e interpretazioni per poter mantenere vivo il legame con la comunità ed espletare al massimo la funzione del museo. Queste nuove tecnologie possono facilitare tale comunicazione, rendendola quotidiana e sempre presente, perciò sarebbe sbagliato non provare a coglierle e massimizzarne i risultati per un miglioramento del servizio rivolto alla cittadinanza.

152 Ellenborgren, Falk, Goldman, 2008 pag. 192

153 La guida è reperibile gratuitamente all’indirizzo http://www.sfmoma.org/explore/multimedia/interactive_features/3

154 http://archives.icom.museum/hist_def_eng.html

Bibliografia

- R. Atkinson, *What's the point of museum websites?*, 16 aprile 2013, <http://www.museumsassociation.org/museums-journal-blog/16042013-whats-the-point-of-museum-websites>, (28/09/2013)
- J. Beel, B. Gipp, E. Wilde, *Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar and Co.*, Journal of Scholarly Publishing. pp. 176–190, 18 aprile 2010 , <http://www.sciplore.org/publications/2010-ASEO--preprint.pdf> , (25/06/2013)
- J. DeMers, *Google's Penguin 2.0 Algorithm: The Definitive Guide*, 8 luglio 2013, <http://searchenginewatch.com/article/2279845/Googles-Penguin-2.0-Algorithm-The-Definitive-Guide>, (18/08/2013)
- K. Ellenbogen, J. Falk, K.H. Goldman, *Understanding the motivations of museum audiences*, contenuto in *Museum Informatics*, a cura di P.F. Marty e K. Burton Jones, New York, Routledge, 2008, pp. 187-194
- E. Everhart, *How to Make Your Keywords Fit Your Marketing Messaging*, 23 giugno 2013, <http://searchenginewatch.com/article/2277087/How-to-Make-Your-Keywords-Fit-Your-Marketing-Messaging>, (25/06/2013)
- A. Goodman, *An end to Metatags Pt. 1*, 2 settembre 2002, <http://www.traffick.com/article.asp?aID=102> , (08/07/2013)
- A. Goodman, *Google Uses Meta Tags Sparingly, But Should You?*, 16 settembre 2002, <http://www.traffick.com/article.asp?aID=105>, (08/07/2013)
- J. Lee, *No. 1 Position in Google Gets 33% of Search Traffic [Study]*, 20 giugno 2013, <http://searchenginewatch.com/article/2276184/No.-1-Position-in-Google-Gets-33-of-Search-Traffic-Study> , (25/06/2013)
- M. Lupi, *Motori di ricerca e visibilità sul Web: come utilizzare al meglio i motori di ricerca nelle attività di web marketing*, Milano, Apogeo 2001
- B. Male, *10 Basic SEO Tips To Get You Started*, 13 gennaio 2010, <http://www.businessinsider.com/10-basic-seo-tips-everyone-should-know-2010-1?op=1> , (11/07/2013)
- M. Maltraversi, *SEO e SEM: Guida avanzata al Web Marketing*, Milano, FAG 2012
- M. Miller, *53% of Organic Search Clicks Go to First Link*, 10 ottobre 2012, <http://searchenginewatch.com/article/2215868/53-of-Organic-Search-Clicks-Go-to-First-Link-Study>, (25/06/2013)
- R. Miller, *Google Unveils Its Cointainer Data Center*, 1 aprile 2009, <http://www.datacenterknowledge.com/archives/2009/04/01/google-unveils-its-container-data-center/> , (17/07/2009)
- R. Miller, *Report: Google Uses About 900,000 Servers*, 1 agosto 2011,

<http://www.datacenterknowledge.com/archives/2011/08/01/report-google-uses-about-900000-servers/> , (17/07/2013)

S. Sancassani, F. Brambilla, P. Marengi, S. Menon, *e-Collaboration : il senso della rete: metodi e strumenti per la collaborazione online*, Milano, Apogeo, 2011

K. Schachinger, *How to Write Title Tags For Search Engine Optimization*, 31 dicembre 2012, <http://searchenginewatch.com/article/2154469/How-to-Write-Title-Tags-For-Search-Engine-Optimization>, (07/07/2013)

K. Schachinger, *How to use HTML Meta Tags*, 1 maggio 2012, <http://searchenginewatch.com/article/2067564/How-To-Use-HTML-Meta-Tags>, (07/07/2013)

N. Simon, *A simple argument for why museum and cultural institutions should care about social media*, 13 aprile 2009, <http://museumtwo.blogspot.it/2009/04/simple-argument-for-why-museums-and.html>, (10/09/2013)

J. Slegg, *New Google Panda Update Rolling Out Now: What Changes Are Webmasters Seeing?*, 19 luglio 2013, <http://searchenginewatch.com/article/2283724/New-Google-Panda-Update-Rolling-Out-Now-What-Changes-Are-Webmasters-Seeing>, (12/08/2013)

D. Sullivan, *Death of a Meta Tag*, 30 settembre 2002, <http://searchenginewatch.com/article/2066825/Death-Of-A-Meta-Tag>, (08/07/2013)

G. Taverniti, *SEO power: strategie e strumenti per essere visibili online*, Milano, Hoepli 2011

L. Toscano, *SEO Strategy*, Trento, UNI Service 2009

D. Vasta, *Web Analytics: le metriche, gli strumenti, le analisi*, Milano, Apogeo 2009

P. Zarokian, *Insights From the Recent Penguin & Panda Updates*, 11 maggio 2012, <http://searchenginewatch.com/article/2174011/Insights-From-the-Recent-Penguin-Panda-Updates>, (18/08/2013)

Sitografia

http://www.googlerank.it/Guide_SEO/Fattori_Ranking_Interni/Tag-Title.asp (28/06/2013)

<http://romagna.info-alberghi.com/rimini/infografica-turismo-riviera-rimini-2012> (10/07/2013)

<https://www.google.com/webmasters/tools/home?hl=it> (11/07/2013)

<http://www.museicomunalirimini.it/> (20/05/2013)

<http://www.dmoz.org/> (13/07/2013)

<http://www.istat.it> (05/08/2013)