

Lsa e semantica applicazioni reali

In questo articolo saranno trattati i seguenti argomenti.

- 1 - Premesse
- 2 - Come funzionano gli algoritmi semantici
- 3 - Limitazioni degli algoritmi semantici
- 4 - Quando Google ne fa uso
- 5 - Utilizzo della semantica nella SEO
- 6 - Cosa aspettarsi dalla SEO Semantica

PREMESSA

Gli algoritmi definiti "semantici" hanno lo scopo di creare un modello matematico in grado di individuare un punteggio di pertinenza di un determinato documento per qualsiasi query di ricerca dell'utente.

Generalmente tali algoritmi vengono utilizzati per la ricerca in documenti riguardanti un settore specifico fornendo funzionalità di ricerca verticale su un settore con un numero ridotto di argomenti/topic.

Ad esempio l'algoritmo semantico LDA è stato utilizzato per la ricerca nell'archivio brevetti degli Stati Uniti.

Algoritmi come LDA sono estremamente pesanti in termini di elaborazione in quanto necessitano di enormi quantità di memoria per elaborare rapidamente ogni termine presente in ogni frase di ogni documento e stabilirne l'affinità con

un topic, motivo per cui ad oggi esistono solo alcuni motori di ricerca sperimentali, e di piccole dimensioni, basati su queste tecnologie.

Il web semantico (lo avrete sentito in inglese Semantic Web) non ha nulla a che vedere con i testi, è tutto basato sui dati (linked data) e sulla loro interpretazione per poi avere la possibilità di predire infinite risposte celate nei dati. In questo libro analizzeremo sia la Semantica dei testi che il semantic web ed i metodi pratici che potrebbero portare dei vantaggi ai nostri siti web.

COME FUNZIONANO GLI ALGORITMI SEMANTICI

Tra i più noti algoritmi semantici c'è l'algoritmo LDA perché in molti lo hanno candidato come possibile algoritmo utilizzato in qualche ranking factor di Google.

Secondo me LDA viene ampiamente utilizzato da Google per la categorizzazione dei documenti sul web per poterli organizzare in gruppi di Topic e ridurre quindi l'insieme di documenti su cui applicare la ricerca dell'utente utilizzando i topic come chiave parziale dell'indice.

L'algoritmo LDA ha in input una query di ricerca e risponde con un valore numerico compreso tra 0 ed un valore massimo (diciamo 1 ma è irrilevante), che indica quanto sia pertinente il documento per la query effettuata.

Non entro nei dettagli matematici perché non ne sarei minimamente capace ma voglio spiegarvi concettualmente cosa accade quando un algoritmo LDA elabora un testo.

Prima di tutto, per poter elaborare documenti ed individuarne la pertinenza per topic specifici, l'algoritmo ha la necessità di conoscere a monte l'insieme di termini che rappresentano ogni singolo TOPIC.

Ad esempio, il topic "calcio" potrebbe essere rappresentato dall'insieme di termini {pallone, calciatore, portiere, dischetto, arbitro, rigore, punizione, ecc...} in cui ogni termine ha un peso diverso. In questo insieme avrà peso elevato "calciatore" e peso basso "dischetto" in quanto il termine potrebbe essere ambiguo ed avere significati diversi in contesti diversi (polisemia). Il motore LDA è in grado di capire che parliamo di calcio anche in frasi poco chiare come "viene messo giù ma non è punizione, si va al dischetto" dove non si fa un riferimento diretto al termine "calcio".

Ovviamente la frase avrà uno score inferiore della frase "il calciatore viene messo giù dal portiere, l'arbitro assegna il rigore, si va al dischetto"

Solo avendo informazioni precise sui topic, diciamo dati di configurazione, LDA potrà fornire un indice di pertinenza di un documento per argomenti specifici.

L'algoritmo non è in grado di scoprire nuovi TOPIC a meno che non venga effettuata una fase di "learning" in cui vengono fatti elaborare solo documenti di un determinato argomento avviando una fase di "topic modeling" ossia in grado di individuare una lista di termini, con il loro peso, che definiscono il concetto/topic. Se il motore LDA continuasse ad apprendere anche in fase di produzione rischierebbe di degradare la qualità della "topic definition", motivo per cui è sempre preferibile creare il "word space" per un topic partendo da documenti attendibili come ad esempio pagine di wikipedia su argomenti specifici.

Qualche anno fa ho realizzato un mio motore di analisi semantica LSA (non di ricerca) provando ad individuare i topic espressi in un documento ed ho constatato empiricamente che tale approccio funziona perfettamente in alcuni ambiti abbastanza lineari, come la medicina, e molto meno in altri dove c'è "maggiore libertà di espressione" dal punto di vista linguistico.

Vi faccio un esempio dell'output del mio algoritmo. Tempo fa avevo un cliente che operava nel settore paramedicale per il trattamento della fimosi ed i miei copywriter avrebbero dovuto studiare a fondo ogni aspetto dell'argomento per poter poi produrre i testi per le pubblicazioni.

Provando a dare in input "fimosi" al mio algoritmo ho avuto come output:

"Problema a scoprire il glande"

"Difficoltà ad avere rapporti sessuali"

"Dolore al prepuzio"

"Cura fimosi serrata"

"Rimedi fimosi non serrata"

"Avere disagio psicologico"

"Ricorrere alla chirurgia"

E tante altre frasi che hanno espresso interamente il concetto e trovate tutte le azioni che rappresentano il topic, non solo i termini ma vere e proprie triple che spiegano soggetto->predicato->oggetto

Veniamo ora all'algoritmo LDA e scopriamo come opera.

FASE 1: Analisi del testo

Inizialmente viene creata una matrice (immaginate un foglio excel) che come righe ha le frasi del documento e come colonne i termini individuati. Nelle celle c'è la frequenza con cui compare il termine nella rispettiva frase (term-document matrix).

Successivamente vengono create nuove matrici topic-document e term-topic che forniranno all'algoritmo tutte le informazioni necessarie per applicare i complessi calcoli matematici. Sulla base dei dati costruiti per il documento il

motore avrà la possibilità di dirci quanto una query sia pertinente con il testo analizzato.

Vi faccio un esempio, se la frase fosse:

"La bimba è stata ripresa dalla maestra ed è diventata rossa come un peperone" la frase sarebbe rilevante per i seguenti TOPIC ma con punteggi diversi:

- Infanzia (bimba, maestra)
- Istruzione (bimba, maestra)
- Cinema (ripresa)
- Cibo (peperone)

Per giungere a questa conclusione il motore LDA dovrà, come detto in precedenza, conoscere tutti i termini legati ai TOPIC specifici.

Per costruire la lista di termini di ogni TOPIC generalmente, in fase di pre lancio, si "danno in pasto" al motore SOLO pagine di un determinato settore in modo da fargli creare la lista dei termini che identificano il topic, poi si procede allo stesso modo per altri settori (TOPIC)

FASE 2: Pulire e semplificare

Lo spazio dei termini e le matrici create nella prima fase di elaborazione sono praticamente ingestibili perché ricche di "rumore", ossia di termini di uso comune o scarsamente rilevanti. In pratica se in un documento scriviamo "pollo" riferito ad una persona ed al suo comportamento ma lo facciamo in un contesto dove il linguaggio è molto formale, la parola "pollo" sarà vista come rumore e tagliata via in quanto usata solo nel nostro documento o in pochissimi altri e non sarà presente nell'insieme di termini che rappresentano il TOPIC.

Dopo la pulizia il motore LDA avrà una matrice molto più snella, precisa e gestibile da interrogare.

Si giunge quindi alla matrice definitiva in grado di rispondere immediatamente alle query. Il peso dell'algoritmo è tutto nella fase di elaborazione delle matrici.

LIMITAZIONI DEGLI ALGORITMI SEMANTICI

La limitazione più grande degli algoritmi semantici è quella di fornire in uscita risultati non sempre precisi a causa di problemi di polisemia (parole che hanno significato diverso in contesti diversi) che alterano in maniera inaccettabile l'output in molti casi.

Altra limitazione che non ne permette l'utilizzo pienamente soddisfacente in ambienti con troppi topic, tipo il web, è la sorgente dell'informazione ricca di rumore.

Le persone producono pagine web scrivendo quello che vogliono, come vogliono, quando vogliono ed arricchiscono le pagine di aneddoti off topic, racconti personali, commenti dei lettori, snippet di testo di articoli correlati che sporcano il contenuto, il tutto porta ad avere un documento ricco di testo extra che ne diminuisce la rilevanza anche se, magari, il contenuto è il migliore in assoluto.

Anche alterando la matrice inoculando termini, nel caso improbabile in cui si sia avuto accesso a tutti i documenti del web, non si avrebbe un giovamento SEO in quanto tutti i documenti presenterebbero il termine allo stesso modo ed un termine per fare la differenza dovrebbe essere il più rilevante del wordspace del topic.

Ad ogni modo queste problematiche sono notevolmente ridotte quando tali algoritmi lavorano su settori verticali infatti è l'unico caso riconosciuto in cui tali algoritmi vengono esplicitamente indicati come algoritmo base del motore. Eccovi alcuni esempi di motori che utilizzano un algoritmo semantico, non a caso sono tutti motori di ricerca verticali.

- vLex, motore di ricerca legale
- Yummly, motore di ricerca settore food
- Thinkglue, ricerca di video
- Zaptravel, ricerca settore viaggi
- Kosmix, ricerca nei social media

QUANDO GOOGLE NE FA USO

Prima di tutto bisogna precisare che lo scopo di un algoritmo semantico, in un contesto come quello del websearch, è quello di capire l'intenzione di ricerca e fornire i risultati più soddisfacenti e vicini al query intent dell'utente.

Altra precisazione da fare è che le query non sono sempre uguali, possono essere distinte in almeno due search intent diversi:

1 - Search intent "navigational" in cui chi effettua la ricerca vuole scoprire, imparare qualcosa, quindi non esiste una risposta precisa ed è necessario fornire una lista di "best guess" che l'utente potrà navigare ed approfondire. Questa è la categoria principale dove noi SEO operiamo e proviamo a posizionare i nostri siti web. In questo ambito potrebbero essere utilizzati algoritmi semantici per categorizzare i contenuti e generare uno score di pertinenza sui search intent. Unica implementazione confermata da Google è stata Hummingbird che ha come scopo quello di capire le intenzioni di ricerca

degli utenti e di rispondere accuratamente indipendentemente dalla forma in cui viene posta la domanda e potrebbe fornire risposte che non contengono esplicitamente la keyword esatta nel testo. Questo algoritmo ha portato un grande cambiamento alle SERP ed alla SEO orientata al Copy in quanto a più domande diverse, ma con lo stesso search intent, Google risponde con SERP molto ma molto simili tagliando quindi fuori milioni di pagine web che in passato erano state create dai webmaster per rispondere a più variazioni della stessa domanda. In conclusione, hummingbird ha tolto valore alle pagine ridondanti che puntavano ad ottenere traffico di long tail e lo ha dato a quelle ricche di "search intent" raggruppati, quindi ai contenuti più convincenti.

Gran parte di hummingbird riguarda la query e la sua comprensione studiandone il contesto di ricerca come ad esempio il dispositivo, il luogo e magari anche il momento in cui viene fatta la ricerca, quindi è un primo passo verso la ricerca semantica ma non come avviene negli algoritmi LDA, in questo caso Google prova a capire "cosa desideri l'utente" non tanto quello che dicano i documenti che sceglierà come risultato.

Dalla mia esperienza personale posso dire che , secondo me, hummingbird si interpone tra l'utente ed il motore classico di Google facendo da interprete intelligente di query traducendole nelle query vecchio stile come ad esempio "pizzeria roma" che poi saranno processate dal layer più basso e maturo del motore.

Ad esempio, query "conversational" come :

- dove mangiare una pizza a Roma
- buona pizzeria a Roma
- chi fa una buona pizza a Roma

Oppure query simili come:

- pizza a Roma
- pizzerie a Roma
- pizza a taglio a Roma

Potrebbero essere semplificate notevolmente da hummingbird in una query "pizzerie+Roma" e fornire gli stessissimi risultati di ricerca in quanto celano lo stesso "search intent" che è quello di pizza+roma

Il resto del lavoro di Google lo farà normalmente il layer sottostante.

Per quanto riguarda il vero semantic web Google ha fatto passi avanti importanti creando il progetto Schema.org con Bing e Yahoo per preparare il motore a linked data.

2 - Search intent di tipo "research" in cui l'utente chiede al motore di ricerca informazioni specifiche su un'entità nota al motore. È questo l'unico caso in cui Google afferma di fare uso di algoritmi semantici e molto probabilmente (quasi sicuramente) non si basa su LDA o simili ma su RDF Traversal, una metodologia che permette di identificare la rilevanza di un concetto analizzandone la vicinanza ad altri correlati e quindi fortemente rilevanti. Questo argomento lo tratterò in un nuovo post specifico in quanto è l'unico approccio al web semantico che per me abbia un senso e di cui ne abbia testato a fondo l'efficienza.

Quando si effettuano query del tipo "research" Google risponde con uno o più Knowledge Boxes i cui dati vengono estratti dal knowledge graph, quindi l'applicazione della ricerca semantica non riguarda le SERP organiche ma solo i side boxes.

Lo scopo è quello di capire il search intent e fornire direttamente la risposta e non una lista di documenti da navigare ed approfondire. Questa applicazione

della semantica in Google riguarda esclusivamente l'utilizzo di linked data e schema.org, altro argomento su cui scriverò un post dettagliato in seguito.

Tirando delle conclusioni direi che Google utilizzi fortemente la semantica per capire cosa voglia esattamente l'utente (query analysis) mentre nella fase di recupero delle informazioni (IR) ne fa uso marginale solo per categorizzare i documenti e restituire più velocemente i risultati (elaborandone di meno avendoli indicizzati) che poi è uno dei punti di forza di hummingbird, veloce e preciso. Attenzione, nell'ultima frase ho utilizzato il termine "indicizzati" volutamente, è un esempio concreto di polisemia e credo che molti di voi lo abbiano inteso nel contesto "motori di ricerca" mentre io intendevo indicizzati nel contesto "database", quindi un processo per avere più velocemente risposte. Questi sono i casi in cui si capiscono i limiti di tali algoritmi. Per poter capire il senso della frase l'algoritmo deve conoscere il topic "motori di ricerca" ma anche il "topic database" e scegliere di conseguenza. In questo caso si può mettere in crisi anche un umano se non conosce il topic "database".

UTILIZZO DELLA SEMANTICA NELLA SEO

Bene eccoci arrivati al dunque, all'argomento che sicuramente vi interesserà di più, ossia se esista o meno la SEO Semantica.

Ebbene sì esiste! Se esiste la SEO e Google in parte fa uso della semantica allora vuol dire che ci sarà pure qualcosa che il SEO può fare a riguardo per piacere di più a Google. Per capire cosa possiamo fare in ambito semantico per ottimizzare un sito dobbiamo vedere il termine SEO in maniera diversa da come siamo abituati a vederla, soprattutto perché anche la figura professionale del SEO è mutata notevolmente. Prima di vedere cosa possiamo fare con la semantica voglio prima esporvi la classificazione dei

professionisti SEO che ritengo più attendibile in modo da poter poi entrare nello specifico del loro ambito d'azione in campo semantico.

1) SEO/SEM esperti in keyword research

Non molto addentrati in problematiche tecniche ma con capacità strategiche e grande abilità nel definire la linea editoriale di un sito web.

2) SEO tecnici esperti di protocolli di rete, linguaggi web e con grande conoscenza dei motori di ricerca e degli algoritmi.

3) SEO Copywriter con solo un'infarinatura di conoscenze tecniche ma abili nel lavorare sui singoli testi per migliorarne il posizionamento per keyword specifiche.

Io, come tanti altri colleghi, parto dalla seconda categoria ma negli ultimi 8 anni ho acquisito piena padronanza anche negli altri due ambiti e la mia esperienza mi porta a dire che nella SEO ogni aspetto sia importante solo che spesso, per un ipotetico cliente, è difficile avere budget necessario per poter curare il proprio sito web a 360 gradi come andrebbe fatto. Vediamo ora ogni categoria di SEO cosa può fare in ambito semantico.

1) SEO/SEM

In fase progettuale di un sito web è fondamentale fare keyword research per decidere a monte dove vogliamo andare a competere e cosa speriamo di ottenere con il nostro lavoro. Il SEM generalmente individua il potenziale del settore, le TOP keyword dove competere e tutta la Long Tail da conquistare. Con l'arrivo di hummingbird è cambiato completamente il concetto di long tail per cui è cambiato di conseguenza anche il lavoro del SEM che dovrà ora fare

liste di search intent (e non più di keyword) che punteranno ad un contenuto specifico da creare, poi per ogni singolo contenuto il copy svilupperà mappe semantiche di concetti per tematizzarla sui search intent specifici, di questo parlo nel punto 3.

2) SEO tecnici

Allo stato attuale delle cose l'unico modo possibile e concreto per referenziare entità (cose, persone, brand), topic (argomenti), meanings (significati) , facts (azioni, notizie, eventi) è quello di utilizzare linguaggi come RDFa o Json-LD ed il vocabolario schema.org per mappare tutte le informazioni riguardanti entità specifiche ed essere certi che Google ne venga a conoscenza. In questo ambito ho fatto molti esperimenti, sono riuscito ad entrare come entity nel knowledge graph e testare le capacità di deduzione di informazioni per inferenza da parte di Google utilizzando RDF Traversal. In piena sincerità, non perché appartengo maggiormente a questa categoria, ritengo che la SEO semantica si faccia solo in questo modo anche se non denigro affatto le migliorie sui testi che possa fare un Copy rafforzando la pertinenza del testo per determinati topic e intent.

Non posso dilungarmi troppo sugli argomenti tecnici della SEO Semantica altrimenti ne uscirebbero almeno un centinaio di pagine ma nei prossimi mesi terrò vari corsi specifici e scriverò anche qualche post a riguardo, preferisco approfondire in ambito copy dove poche nozioni potrebbero esservi utili già nell'immediato per prendervi qualche piccola soddisfazione su keyword poco competitive.

3) SEO Copy

Considerando i grandi cambiamenti di Google per quanto riguarda l'interpretazione delle query di ricerca e l'identificazione del search intent è

ovvio che sia cambiato qualcosa anche nella strategia di composizione dei testi. Lavorare all'ottimizzazione dei testi è spesso un lavoro, che in una strategia SEO, ha un costo molto più elevato di quanto non si possa rientrare economicamente di conseguenza, quindi lo vedo come un task secondario rispetto alle ottimizzazioni concettuali e tecniche onsite che possono cambiare completamente la vita di un sito e far monetizzare in tempi ragionevoli.

Ciò nonostante ritengo che sia di indubbia utilità una strategia volta all'adeguamento graduale dei contenuti per rispondere in maniera più dettagliata alle intenzioni di ricerca degli utenti, svolgendo il lavoro sul lungo termine si diluisce l'investimento e si ottengono risultati gradualmente. Ma come può fare un SEO Copywriter a migliorare il posizionamento di una pagina web utilizzando la semantica?

Se consideriamo che l'algoritmo LDA è il più plausibile bisognerebbe fare reverse engineering e produrre documenti esattamente come se li aspetta l'algoritmo per un determinato topic. Partiamo dal vedere "cosa si aspetta" e poi arriveremo al come. L'algoritmo LDA partendo da topic noti riesce a capire in che percentuale ogni documento (e singola frase) sia pertinente con essi, quindi ogni documento sarà indubbiamente una ricetta composta da percentuali ipotetiche come 30% sport, 20% food, 20% tv e 30% amicizia se ad esempio il documento parla di una serata tra amici a vedere la partita in salotto mangiando pizza e bevendo birra.

Solo questa ricetta di topic, più o meno vicina a quella ottimale, dà la certezza al motore che stiamo parlando di un argomento ben preciso e non di tutte le migliaia di sfumature possibili.

Veniamo al come. Come può fare un copywriter a capire quali siano i topic da utilizzare, in che percentuale e con quale bilanciamento? In SEO Cube utilizziamo "Relevance" un tool realizzato da me ma ho fatto anche una versione semplificata sul web nella nostra SEO Suite SEOZoom.

Al momento in SEOZoom, essendo ancora in fase di sviluppo, ho semplicemente calcolato una matrice TF (term-frequency) che indica (non rabbrivite) la keyword density di ogni termine nell'insieme di documenti. Spero sia chiaro che lo scopo non è quello di usare le KD come fattore di ranking ma individuare i termini più comuni nei documenti posizionati per una keyword. La fase di sviluppo successiva sarà quella di procedere nello sviluppo passando al metodo tf-idf che significa Term Frequency - Inverse Document Frequency, che ha lo scopo di dare maggiore importanza ai termini che compaiono nel documento, ma che in generale sono poco frequenti e quindi non facenti parte del linguaggio comune ma che invece rappresentano "il succo" dell'articolo.

Sembrerebbe che per ottimizzare i testi sia assolutamente necessario un software specializzato ma non è difficile fare un buon lavoro anche a mano realizzando delle semplici mappe semantiche come si fa fare ai ragazzini alle scuole superiori per allenarsi nell'analisi (noi ad informatica lo facevamo).

Una mappa semantica è estremamente semplice da realizzare, basta prendere un foglio A4, scriverci l'argomento principale nel centro in un cerchietto e poi creare tutti i concetti collegati che vi vengono in mente intorno collegandoli con una linea. Anche i concetti secondari potete approfondirli con sotto-concetti. Alla fine avrete una mappa degli argomenti rilevanti da trattare, non bisogna essere dei geni, dovete solo conoscere quello di cui state parlando.

In questo modo, senza alcuna conoscenza matematica o informatica potrete essere rilevanti e scrivere contenuti migliori.

In un progetto SEO ci sono aspetti prioritari come quelli tecnici (salute del sito), ed alcuni secondari, ma non poco importanti, come quelli strategici (conoscenza ed obiettivi) e quelli qualitativi (contenuti e usabilità), come nel

mondo reale, se vuoi competere, ti metti in perfetta forma, ti alleni e studi, migliori aspetti specifici e curi i dettagli e poi , forse, avrai qualche possibilità di vincere. Se curi solo un aspetto, trascurandone altri, sarai come me che sorseggio birra sulla spiaggia con un po' di pancetta di troppo e di certo non l'atleta vincente a cui vorremmo sempre paragonare i nostri siti web.

COSA ASPETTARSI DALLA SEO SEMANTICA

Anche la risposta a questa domanda va necessariamente divisa in tre risposte specifiche per ogni tipologia di attività di ottimizzazione che ho elencato in precedenza, SEM, Tecnica, Copy.

1) SEM, keyword intent research

Da questa attività mi aspetto il risultato più visibile in tempi brevi in quanto si ha la possibilità di moltiplicare le possibilità di fare visite per ogni singola pagina che sarà prodotta o ottimizzata nel sito.

Da una buona organizzazione derivante da questa attività potrebbe essere completamente rivalutato un sito web da parte di Google quindi mi sento di darci importanza prioritaria. C'è da dire però che nel settore tutti i SEM che conosco ancora non hanno intrapreso questa strada e lavorano ancora su enormi keyword list per far produrre articoli specifici per la long tail.

2) SEO Semantica

Questo è l'aspetto che ritengo più interessante in quanto misurabile e gestibile al 100% utilizzando i linguaggi ed i vocabolari scelti da Google, quindi dicendogli esattamente quello che vogliamo parlando la sua lingua ed essendo certi che comprenderà il nostro messaggio.

Con RDF, Json-Ld, schema.org possiamo dire a Google qualsiasi cosa su qualsiasi argomento essendo certi che ci capirà e che probabilmente in un futuro non troppo lontano potrebbe fidarsi fortemente di noi e consigliarci tra i risultati di ricerca anche per query di tipo "research" ossia nei knowledge boxes. La cosa più bella di queste metodologie è che puoi costruire l'informazione di Google mattone dopo mattone essendo certo che lui li metterà tutti insieme (merge) ottenendo la piena conoscenza di qualsiasi topic o entità specifica. Conoscenza ancora più approfondita di quanta non gliene abbiamo fornita noi in quanto, grazie alle relazioni tra entità ed alle ontologie (significati) sarà in grado di dedurre anche tutto quello che non gli abbiamo espressamente insegnato (RDF Traversal).

Dalla SEO Semantica tecnica mi aspetto il massimo che si possa ottenere in quanto è quello che si aspetta e richiede Google dai siti web moderni.

3) SEO Copywriting

La scrittura di testi intent oriented ritengo che sia fondamentale per non sprecare denaro e lavoro senza ottenere nulla. Nel settore editoria, dove lavoro tanto da circa sei anni, si vedono degli sprechi di risorse enormi, realtà con decine e decine di giornalisti che producono articoli di indubbia qualità ma che non ottengono assolutamente nulla nelle ricerche organiche. Il motivo principale è che il giornalista o il Copy lavora come se l'articolo dovesse andare in stampa su cartaceo dove il lettore già lo hai e tutto quello che devi fare è attirare la sua attenzione con un titolo ed una intro ad effetto, tutte cose che nel web proprio non funzionano. Quando faccio formazione ai copy delle aziende con cui collaboro mi viene chiesto spesso un metodo, una formula per fare sempre centro con ogni articolo ma ovviamente non esiste. Esiste però una semplice regola che ti permette di ottenere decisamente di più, quella di "non divagare" e di non riempire la pagina perché bisogna

raggiungere un certo numero di battute. Ogni argomento si esaurisce con più o meno testo e per alcuni ci sta anche bene il misero articoletto da 300 parole. Veniamo alla pratica, che alla fine conta più di 1000 parole, facendo una breve checklist dei task fondamentali per scrivere un articolo semanticamente ottimizzato.

- 1) Effettuate la ricerca che vi interessa su Google.
- 2) Stampate su carta e leggete i primi venti articoli proposti da Google e contemporaneamente create la mappa semantica aggiungendo degli elementi ogni qualvolta identificate qualcosa di intopic che vi era sfuggita.
- 3) Evidenziate sui fogli le triple soggetto->predicato->oggetto più ricorrenti, vi serviranno perché ne farete uso nel vostro testo originale. Ad esempio "il nuovo SISTEMA OPERATIVO -> RIDUCE IL CONSUMO -> della BATTERIA", sono queste le possibili triple RDF che Google potrebbe riuscire ad estrarre in maniera implicita dal testo.
- 4) Se negli articoli individuate dei concetti marginali come ad esempio BIRRA e PATATINE quando si parla di partita in TV non ignorateli ma segnateli con un altro colore, rappresentano i TOPIC secondari (ingredienti) della vostra ricetta semantica.
- 5) Ora avete la bozza della vostra mappa semantica, ridisegnatela "in bella" mettendo di lato ad ogni argomento una percentuale che ne indichi il peso. Partite dal centro mettendo 100% all'argomento principale poi seguite ogni linea verso i sotto argomenti ed assegnateci una percentuale facendo in modo che la somma dei sotto argomenti di ogni nodo arrivi a 100%.
- 6) Scrivete il vostro articolo facendo in modo che il numero di frasi su ogni argomento, rispetto al totale, rientri più o meno nella percentuale designata. Non è complicato sono semplici percentuali, su 10 frasi totali, se un argomento è al 30% scrivete 3 frasi a tema.

Un approccio del genere è semplice ed alla portata di chiunque in quanto non richiede alcuna skill tecnica, basta essere bravi a scrivere e studiare i testi che hanno risultati migliori nelle SERP.

COSA POSSIAMO ASPETTARCI

In conclusione, cosa dobbiamo aspettarci dalla semantica applicata ai motori di ricerca? Vi dico il mio punto di vista partendo da alcune premesse.

1) Google non ha mai affidato la scelta dei risultati in SERP solo alla capacità di analisi dei contenuti da parte dei suoi algoritmi.

2) Google sa benissimo che noi SEO continueremo a provare a manipolare i suoi risultati qualsiasi cambiamento egli possa mai fare ai suoi algoritmi.

3) Per evitare manipolazioni troppo importanti Google ha delegato la valutazione dei siti da mettere in SERP a più di 200 programmi/moduli/routine in grado di valutare aspetti specifici del sito/contenuto. In questo modo non esiste singolo intervento che da solo ci possa permettere di prevaricare gli altri siti in SERP.

4) Google Panda da solo ha ulteriori 500 segnali utili alla valutazione dei contenuti, dell'autorità di siti ed autori e dell'autenticità dei testi.

5) Quando si manipola un testo per posizionarsi meglio , questo può accadere solo ed unicamente se i siti competitor sono fortemente carenti sugli altri 199 fattori o almeno alla vostra pari e non soddisfino meglio di voi i 500 segnali e

più di Panda, di certo non supererete Wikipedia con un testo ottimizzato meglio. Con questo intendo non meglio di come era prima, nemmeno se lo fate meglio di Wikipedia stesso.

Volendo semplificare il tutto, giusto per rendere l'idea a chi non è un tecnico, immaginate che la somma massima delle 200+ valutazioni sia 100% e che ogni fattore del posizionamento abbia un peso più o meno consistente. Se lavorate su un solo aspetto dei 200 di certo potreste migliorare il ranking ma solo se per la keyword specifica l'ago della bilancia tra i siti in SERP era proprio il fattore contenutistico, non funzionerà sempre.

A mio avviso, ritornando al discorso principale, la SEO Semantica vera, quella fatta di linked data può fare una differenza enorme e dare un impatto incredibile su molti dei 200+ fattori di ranking e addirittura rappresentare la soluzione ideale se siete stati penalizzati da Panda. In che modo? Ve lo spiego subito.

Google ad oggi conosce miliardi di siti web, alcuni li reputa più autorevoli di altri in quanto siti ufficiali di brand importanti o siti super famosi come wikipedia, facebook ecc... La popolarità di un sito web, calcolata in qualche modo da Google, che noi tendiamo a chiamare Trust Rank oppure, sfruttando metriche di terzi, Domain Authority, è qualcosa che non va intesa come semplice metrica di un sito web ma anche del BRAND (entità) che ne detiene la proprietà. Questi siti, diciamo TRUSTED, godono di maggiore attenzione da parte di Google e spesso vengono posizionati anche con contenuti miseri in quanto per Google è quasi certo che siano meglio di quelli degli altri ancor prima di perdersi del tempo per analizzarli a fondo, oppure, caso più attendibile, questi contenuti, anche se miseri, inquadrati nel contesto

semantico del sito intero lo completano e lo arricchiscono fornendo valore al sito ed al lettore.

Negli ultimi due anni ho lavorato molto su questo tipo di informazioni ed ho provato a far crescere l'autorevolezza di vari siti web con campagne di link building semantiche brand oriented. L'impatto sulla crescita organica è stato notevole e non ho avuto bisogno di puntare a keyword specifiche ed anchor forzate è bastato lavorare sul CONTESTO in cui i link venivano piazzati, un contesto in grado di far capire a Google che il nostro sito web non è forte per una keyword ma per tutti i search intent che mi interessavano, un solo link che ne vale 50 insomma. Nella semantica ogni cosa ha un valore e significato diverso a seconda del contesto in cui viene utilizzata infatti qualche anno fa, quando impazziva il termine "Content is king" ho coniato una frase più vicina alla mia visione "Context is king".

Questo metodo, abbastanza rudimentale in quanto presuppone che Google INTERPRETI quello che scriviamo nelle pagine web, ha un'efficienza ridotta, ma non trascurabile, per vari motivi:

1) L'estrazione implicita delle entità, topic, significati è inefficiente, lenta, imprecisa, pesante quindi richiede più tempo e pazienza per attendere qualche possibile ed incerto risultato. Spesso l'incapacità di Google nell'estrarre entità potrebbe addirittura creare problemi e penalizzazioni Panda non meritate in quanto Google potrebbe credere che stiamo creando "overlapping content" quando invece stiamo solo approfondendo e dettagliando informazioni riguardo la stessa entità.

2) L'estrazione implicita funziona solo su entità note e Google non estrae entità nuove (se non le conosce cosa dovrebbe estrarre?) quindi un nuovo

Brand andrebbe a vuoto vanificando l'intera strategia. Google si limita a scoprire nuovi fatti sulle entità già schedate.

3) Per funzionare bene, anche i siti ospitanti per i link devono avere una propria entità nota ed autorevole altrimenti la loro parola vale zero.

Per me ha funzionato alla grande perché il mio cliente é un Brand famoso nel mondo, con budget adeguato per permettermi di pubblicare su siti autorevoli nel suo settore.

Risultato? 750 keyword portate in prima pagina in un anno e mezzo!

Immaginate quanto sarebbe costata una campagna di link building mirata ad ogni singola keyword oppure un lavoro di revisione/implementazione del testo di ogni singola pagina web.

Nel mio caso specifico potevo facilmente far capire a Google CHI consigliasse CHI e per CHE COSA. Se non si fosse trattato di un top brand Google non avrebbe capito a chi mi riferissi ed avrei dovuto fare link building tradizionale con URL o Anchor esatte.

In alternativa, utilizzando RDF e Schema, abbiamo la possibilità di superare le problematiche legate all'interpretazione del testo, porteremo Google a fare estrazione esplicita di entità ed avremo anche la possibilità di fargliene conoscere di nuove, collegate ad un sito web ed a tante altre migliaia di informazioni che la descrivano appieno. Non c'è limite al tipo di informazione che si può fornire a Google con schema.org seminando "pillole di informazione" in qualsiasi contesto web a tema, Google le raccoglierà tutte e poi ne farà il merge per ricomporre la totalità della conoscenza sull'entità, e cosa più importante, saprà quali sono i siti web che hanno contribuito

maggiormente, e meglio, alla costruzione dei dati, i siti che acquisiranno la sua fiducia (possibilmente trust rank o come preferite chiamarlo).

Con la seconda opzione lo scenario cambia completamente, prima potevamo solo produrre testo e sperare che Google capisse cosa volevamo dire, ora abbiamo la possibilità di parlare la sua stessa lingua, dirgli quello che vogliamo ed avere la certezza che ci capisca e ci valuti adeguatamente. Altro aspetto molto importante è che utilizzando schema.org non solo parlerete la stessa lingua di Google ma anche di Bing, Yahoo! e, cosa non trascurabile se puntate anche al mercato russo, di Yandex.

Se siete ancora scettici a riguardo vi basti pensare che ad oggi, agosto 2015, Google ha nel suo Knowledge Graph 1 miliardo di entità censite (wow ci sono anche io e la mia azienda) , 51 miliardi di notizie (facts) ed il tutto disponibile indipendentemente in 40 lingue (fonte Denny Vrandecic di Google).

Denny Vrandecic è un personaggio autorevole in Google e come lavoro fa l'ontologista, segnale importante che Google stia lavorando sul web semantico e soprattutto sulla definizione di modelli semantici (*) che possano dare senso a tutte le informazioni presenti sul web sotto forma di dati strutturati e non più solo testo.

Non preoccupatevi, dopo questa infarinatura generale sugli argomenti importanti procederò con esempi pratici di come si possa fare SEO Semantica indipendentemente dal tipo di SEO in cui vi immedesimate, la teoria è interessante ma la SEO è azione in primis, quindi vediamo come mettere tutto in pratica.

SEO SEMANTICA IN PRATICA

IMPARARE LA LINGUA DEL PANDA

A parte i 200 fattori del posizionamento, argomento di moda ma che nessuno è in grado di affrontare pragmaticamente, c'è l'algoritmo Panda che ha il compito più arduo, quello di valutare se un sito è autorevole e se i contenuti delle sue pagine siano di qualità per i possibili utenti.

Per capirlo Panda ha la necessità di individuare, prima di tutto, l'argomento centrale della pagina web (entità principale), dopodiché deve capire cosa stiamo dicendo di specifico su quell'argomento.

In siti web molto complessi, con molte pagine che spesso trattano gli stessi argomenti da punti di vista diversi possono verificarsi problemi concreti con Panda perché, invece di vedere le nuove pagine come approfondimento le vede invece come contenuto duplicato o leggermente differente "overlapping content".

Essendo Panda l'algoritmo ufficiale delegato al controllo qualità dei contenuti e dell'informazione racchiusa nelle pagine è proprio con lui che dobbiamo imparare a dialogare, nel modo che preferisce e soprattutto spiegandogli bene cosa intendiamo dire nei nostri documenti per evitare che si possa confondere.

Lo scopo di Panda è quello di aumentare la qualità dei documenti e dei siti in SERP quindi Google dice di non preoccuparsi se le nostre pagine hanno uno standard qualitativo alto ma c'è da dire che un'organizzazione sbagliata dei contenuti, link interni non strutturati bene possono confondere Panda facendogli credere che stiamo provando a manipolare le SERP producendo thin content per "colpire più keyword variations" e che quindi non siamo "degni della top 10".

Iniziamo quindi a conoscere Panda o almeno a capire cosa si aspetti da noi, analizzando i 23 punti delle linee guida fornite da Amit Singhal di Google. Ad ogni punto aggiungerò le mie considerazioni poi in seguito approfondirò

spiegandovi come fare a soddisfare le richieste di Panda utilizzando un approccio semantico.

- Ti fideresti delle informazioni presentate in questo articolo?

La prima domanda presentata da Amit è estremamente generica quindi presuppongo che su di essa siano concentrati numerosissimi dei 500 controlli algoritmici di tipo diverso. A mio avviso i fattori principali e sono:

- 1) la fiducia che Google ha verso il dominio che ospita l'articolo (trusted entity)
- 2) la fiducia che Google ha verso l'autore
- 3) indubbiamente la link popularity dell'articolo può fare la differenza
- 4) le interazioni dei lettori potrebbero essere un segnale importante
- 5) presenza di dati strutturati che facciano chiarezza e scaccino ogni dubbio del motore

- Questo articolo è stato scritto da un esperto o entusiasta che conosce a fondo l'argomento oppure si tratta di un articolo superficiale?

Punto interessantissimo perché evidenzia il punto di vista di Google riguardo l'entità "autore" dell'articolo. Se l'entità non è trusted allora probabilmente l'articolo sarà considerato di default superficiale (fino a prova contraria)

- Il sito ha articoli duplicati, troppo simili o ridondanti sullo stesso argomento o argomenti simili che differiscono solo leggermente con lievi variazioni delle keyword?

Questo punto si riferisce alla vecchia pratica di voler scrivere un articolo specifico per ogni parola chiave anche se solo leggermente differente da

quelle su cui abbiamo già creato un contenuto in precedenza. Panda può credere che degli articoli siano ridondanti quando in realtà non lo sono affatto, in queste circostanze i dati strutturati e schema.org possono evitarci penalizzazioni ed anche fare aumentare il traffico del sito.

- Daresti serenamente la tua carta di credito a questo sito?

Questo aspetto ritengo sia fortemente legato ad un'altra entità, la branded entity (azienda) che detiene la proprietà dell'entità website su cui scrivono le varie entity author. Questo è uno degli aspetti su cui lavoro maggiormente con il semantic web.

- L'articolo presenta errori grammaticali, stilistici o fornisce informazioni errate?

La presenza di errori di scrittura ricorrenti nei vari documenti del sito è giusto che indichino una scarsa qualità, idem per lo stile di presentazione della pagina ma l'aspetto più interessante è quello riguardante la validità delle informazioni della pagina. A riguardo, ancora una volta, entra in gioco il potere del semantic web e nello specifico gli algoritmi di fact extraction. I fatti possono essere estratti e può esserne verificata la correttezza perché si possono confrontare con le stesse informazioni fornite da altri siti, ne avrete sentito parlare come "algoritmo antibufala". Anche in questo caso possiamo chiarire il nostro messaggio utilizzando schema.org inducendo Google a fare estrazione esplicita e non provare ad indovinare

- Gli argomenti vengono trattati in quanto alimentati da un genuino interesse nel fornire informazioni di qualità all'utente oppure sono

generati nel tentativo di intercettare keyword ed ottenere più traffico organico?

Questo punto, per essere verificato, necessita di due verifiche di tipo diverso. Una a livello di sito in cui Google deve verificare che non ci siano contenuti simili che provano a competere per la keyword principale. Uno a livello di pagina in cui Google deve, prima di tutto, capire se c'è sovraottimizzazione o forzature, poi verificare la qualità di quello che abbiamo scritto.

Per quanto riguarda la qualità, ancora una volta possiamo dare qualche dritta a Google utilizzando schema.org e dandogli la possibilità di confrontarci con altre fonti.

- L'articolo fornisce informazioni e contenuti originali, report e ricerche originali o analisi originali?

Argomento ormai maturo su cui Google avrà sicuramente numerosi tipi di controlli ma con l'utilizzo di markup semantici possiamo fare molto di più dimostrando da subito a Google che stiamo dicendo qualcosa di nuovo, aolo nostro e di qualità.

- La pagina fornisce un sostanziale valore aggiunto rispetto alle altre pagine presenti nei risultati di ricerca?

Questo è il compito principale di Google, lui analizza e sceglie le pagine che ritiene migliori ma la domanda è molto interessante perché, riflettendoci, con schema.org, utilizzando i giusti markup, possiamo non solo dire a Google di cosa parliamo ma possiamo anche fargli capire quali altre pagine di altri siti web lo facciano. In pratica possiamo usare i markup semantici per dire a

Google che parliamo di un argomento ed indicargli le pagine con cui ci riteniamo all'altezza di competere.

- Quanto controllo qualità viene fatto sul contenuto?

Questo aspetto credo venga valutato pagina per pagina per ogni sito e che poi venga generato un valore di qualità per il sito intero.

La presenza di markup semantici a mio avviso sono un fortissimo segnale di controllo qualità se gestiti bene.

- L'articolo descrive entrambe una storia da entrambi i punti di vista?

Per sapere se l'articolo è esaustivo e ricopra appieno l'argomento e gli eventuali punti di vista differenti bisogna presupporre che Google sia a conoscenza dell'entità principale di cui si parla e di tutti i fatti che la riguardano. Anche questa volta è indiscutibile il fatto che per effettuare questo tipo di controlli Google necessita di avere un database di fatti da interrogare velocemente e in massa, ad ogni pagina nuova da analizzare, dovrebbe confrontarla con tutti i testi esistenti sul web. Anche per questo fattore schema.org può fare nettamente la differenza.

- Il sito viene considerato un'autorità riguardo il suo argomento?

Questa domanda fa capire chiaramente che Google categorizzi i siti web, o sezioni di siti, per topic, il che fa pensare che in uno dei 500 "programmi di controllo" di Panda ce ne sia uno che faccia uso dell'algoritmo LDA o simili per poter generare score di pertinenza dei documenti per tutti i topic noti.

Altro aspetto importante da prendere in considerazione è quello dell' "essere un autorità riguardo il suo argomento", che subito ci porta a pensare ad un algoritmo che controlli e valuti citazioni e menzioni da parte di altri siti autorevoli sullo stesso topic. Questo argomento è un altro su cui lavoro incessantemente per i miei clienti, link building semantica che ho rinominato un "Trust Building".

- I contenuti vengono fatti produrre massivamente da un gran numero di copywriter, o pubblicati su una larga rete di siti, e quindi le pagine singole o siti web non ricevono la giusta attenzione o cura?

Esistono vari modi per fare blogging o editoria online ma i due più ricorrenti sono:

1) creare un sito/portale/blog e curarlo nei minimi dettagli per farlo diventare fonte di informazione di qualità

2) creare reti di blog , sotto lo stesso publisher, e puntare alla quantità degli articoli che vengono affidati a copywriter a pochi euro per articolo

Evidentemente Google predilige il punto 1 ma, dove c'è scarsità di informazioni di qualità, i circuiti di blog riescono a conquistare comunque traffico organico, lo perderanno quando qualcuno affronterà la loro nicchia con l'approccio del punto 1.

Per questo punto specifico basti sapere che Google prova a capirlo, quindi alzate la qualità anche se fate parte del tipo 2.

- L'articolo è stato editato bene, oppure si mostra prodotto in modo approssimativo e frettoloso?

In questo caso Amit sembra riferirsi a come sia stato impaginato, formattato, presentato ed arricchito di media l'articolo. Un articolo curato presenterà sicuramente dei segnali percepibili aliticamente come ad esempio la suddivisione in paragrafi non troppo grandi, la decorazione con tag semantici come strong, em ed un corretto uso degli H1..H5 per chiarire la gerarchia delle informazioni nel documento. In alcuni casi l'utilizzo di liste ordinate OL e non ordinate UL migliora la leggibilità ed evidenzia l'organizzazione di concetti o elementi presenti nella pagina. Anche l'utilizzo di immagini, video e link di approfondimento sono segnali di qualità e cura del contenuto.

L'utilizzo di linked data nell'articolo è un inequivocabile segnale di cura del dettaglio sopra la media in quanto ci teniamo affinché Google capisca chiaramente tutto.

- Per una query di ricerca riguardante il settore della salute, vi fidereste delle informazioni fornite da questo sito?

Il settore salute viene sicuramente trattato in maniera particolare da Google in quanto quasi tutti cercano su Google ancor prima di rivolgersi al proprio medico per un consulto. Infatti proprio il settore medico è uno di quelli con il maggior numero di entità e vocabolari specializzati sul web. Per i siti web di settore ritengo che l'utilizzo di schema.org possa veramente cambiare drasticamente il traffico organico ed aumentare enormemente il trust.

- Riconoscereste questo sito come una fonte autorevole quando viene menzionato per nome?

Altro argomento a me molto vicino, quello del trust propagato ad un sito web dalla Trusted Entity che lo ha creato. Il sito stesso diventa una Branded Entity (prodotto) di proprietà del Brand autore ed acquisisce autorevolezza e Brand Awareness. Di conseguenza il Nome del sito diventa keyword di ricerca su Google e rappresenta, in casi di successo, la principale fonte di traffico organico per il sito. L'utilizzo di linked data fa decisamente la differenza in questo ambito specifico.

- Questo articolo fornisce una descrizione completa ed esauriente dell'argomento?

Molti pensano che per superare questo punto basti scrivere un articolo più lungo del normale, diciamo 3000 parole ,commissionandolo magari ad un copywriter ,ma non è quello il punto. Un articolo per essere completo deve trattare a fondo il topic di cui parla ma deve farlo sul serio ossia toccando ed approfondendo subtopic correlati e ben contestualizzati. Anche in questo caso viene da pensare allo score sui Topic prodotto dall'algorithm LDA che qui ci calza a pennello. Immaginate un alto score LDA accoppiato ad un buon markup semantico che confermi il tutto ed il gioco è fatto. Lo score alto per LDA non saprete mai se lo avete raggiunto, non c'è modo di saperlo, ma il grosso del lavoro potete farlo con schema.org sapendo per certo che avrete comunicato il necessario a Google.

- Questo articolo contiene un'analisi profonda o informazioni interessanti ed è lontano dall'essere ovvio?

Ritengo che per Google capire se in una pagina web ci sia solo del testo buttato lì sperando di fare visite organiche oppure un contenuto con informazioni interessanti che possano dare valore al lettore non sia affatto

difficile a meno che l'estrazione implicita dei concetti e delle entità sia avvenuta senza alcun problema. Anche in questo caso i markup semantici, abbinati ovviamente ai buoni contenuti, possono fare la differenza fornendo a Google le triple RDF (soggetto, predicato, oggetto) che colmino la conoscenza superficiale dell'argomento. Con il testo non è semplicissimo che ciò avvenga ed è probabile che non avvenga affatto.

- È il tipo di pagina web che aggiungerei ai preferiti, condividerei con un amico o consiglieresti?

Quest'altra domanda mi fa pensare a due aspetti che Google potrebbe prendere in considerazione:

1) "Aggiungerei ai preferiti" potrebbe essere monitorato attraverso le visite di ritorno alla stessa pagina

2) "Condividerei con amici" visite provenienti da email/gmail o da social network

3) ancora una volta menzioni e citazioni, magari nei commenti su articoli a tema in altri siti web o forum, per rimanere nell'ambito del suggerimento agli amici.

- Questo articolo ha un numero eccessivo di pubblicità che distraggono o interferiscono con il contenuto principale?

È ovvio che il tentativo di monetizzare a tutti i costi sia una forzatura che sporca l'intenzione dell'articolo stesso. Quando ciò accade è ovvio che Google

percepisca che lo scopo dell'articolo non sia quello di fornire informazioni ai propri lettori ma quello di posizionarsi e trarne profitto. In questo caso bisogna solo fidarsi nel buon senso dei webmaster e soprattutto dei centri media che cannibalizzano completamente l'above the fold dei propri siti partner.

- Vi aspettereste di trovare questo articolo stampato in una rivista, enciclopedia o libro?

Linked data è nato sulla base di informazioni note, create dalla community freebase e poi wikidata (versione di wikipedia in forma strutturata) ma anche appoggiandosi a numerose trasformazioni di libri e studi scientifici e medici in formato database. Non mi stupirebbe se Google confrontasse i dati proposti dalle pagine web riguardanti un Topic specifico con i dati estratti da fonti bibliografiche e quindi altamente credibili sia dal punto di vista dell'autore che del contenuto. Anche in questo caso l'ottimizzazione semantica delle pagine è un dovere, semplifichiamo la vita a Google e saremo ripagati.

- Gli articoli sono brevi, inconsistenti, o carenti in informazioni di utilità?

Il problema del thin content è un problema noto e di vecchia data ma molti hanno confuso l'aggettivo thin (sottile, di poco spessore) con short, quindi articoli troppo brevi ed hanno reagito al problema incrementando la lunghezza in parole dei propri articoli. Non è quello che voleva Google, non è importante che un articolo sia per forza lungo, l'importante è che tratti gli argomenti che vuole esporre nel modo giusto e con informazioni esatte e di qualità.

C'è ancora bisogno che dica che schema.org può risolverci il problema?

Immaginate di scrivere un articolo di medie dimensioni, uno di quelli che può sembrare carente ma che invece è un concentrato di informazioni utili, bene

con i dati strutturati questo concentrato di informazioni sarà svelato con chiarezza a Google.

- Le pagine sono create con grande cura ed attenzione per il dettaglio oppure hanno poca attenzione per i dettagli?

Cosa si intende per cura dei dettagli è da interpretare. Google si riferisce a come curiamo la pagina per l'utente o a come curiamo la pagina per il motore di ricerca? Io credo si riferisca all'usabilità della pagina, al numero di distrazioni intorno all'articolo, al numero di approfondimenti. Questo punto si sovrappone in parte ad un punto precedente in cui parlavo della qualità ed attenzione con cui i contenuti vengono proposti all'utente. Il dettaglio potrebbe consistere, nello specifico, in quanti approfondimenti H2,H3 ecc... vengono forniti al lettore, la profondità del dettaglio insomma.

- Gli utenti si lamentano quando vedono pagine di questo sito?

Da anni si parla di sentiment analysis ed è anche possibile che Google, contestualmente all'individuazione di menzioni e citazioni identifichi anche il tono e la positività o negatività della menzione o citazione. Nel caso specifico credo che Amit si riferisca maggiormente alle DMCA ossia a quante lamentele reali siano pervenute a Google da parte di utenti a cui siano stati copiati contenuti o riguardo altre violazioni del sito.

Abbiamo analizzato punto per punto cosa si aspetta Google dai nostri siti web e dai contenuti che produciamo, ma in che modo possiamo agire praticamente per avere siti e contenuti che possano essere premiati da Google?

Non preoccupatevi, ci arriveremo per gradi, per ora è fondamentale che vi sia chiaro tutto quello che è stato detto fin ora, sono elementi fondamentali da conoscere per avere le idee chiare su quello che vedremo in seguito.

Adesso voglio parlarvi di cosa sia il web semantico per i motori di ricerca entrando più nello specifico cosicché vi possa essere chiaro il perché della sua enorme importanza per la SEO moderna e soprattutto per contrastare la cattiva informazione fatta da chi confonde la semantica linguistica con il web semantico. I motori di ricerca devono comportarsi adeguatamente qualsiasi sia la lingua dei documenti che indicizza. Se dovesse essere scritto un algoritmo che capisca il testo dal punto di vista linguistico, diverso per ogni lingua del mondo, non potrebbe esistere SEO Semantica che funzioni globalmente, con linked data invece questo è possibile, vediamo come funziona il web semantico.

SEMANTIC WEB - COME FUNZIONA