

I gravami dello spam: conseguenze pratiche e soluzioni, un caso concreto

Pochi giorni fa ero a Trieste da un cliente. Come spesso succede ci sono andato con la moto: tempi di percorrenza certi, nessuna coda e relax in una giornata praticamente estiva.

Il fine tuning della moto è una mia ossessione. Tengo sempre controllati i consumi e il gioco è consumare il meno possibile a parità di percorso e di stile di guida.

L'ideale sarebbero gli irraggiungibili 40km/litro del mio primo gilera 124-5 marce giubileo (perché poi giubileo? Mah!) col quale ho attraversato gli anni '70 del secolo scorso.

Mi beavo dei notevoli 22 km/litro dello scarabeo 500 GT dell'andata (autostrada a velocità codice) quando mi è venuto in mente di controllare, prima del ritorno, la pressione delle gomme. Non che ci fossero sintomi particolari, solo un andamento che non ricordavo sui giunti d'asfalto e una leggera rigidità del mezzo quando piegava. Sensazioni quasi impercettibili.

Infatti le ruote erano sgonfie: 1,5 bar a fronte dei 2,2 richiesti nel libretto uso e manutenzione e subito ripristinati.

Detto fatto, i consumi al ritorno schizzano a 25,9 km/litro.

Bel colpo: oltre il 17% in più senza toccare niente, a parte la pressione delle gomme.

Bel colpo e scemo io a non controllare con più metodo il mezzo, o meglio, a concentrare i controlli solo sulle parti meccaniche, sul motore, su quello che spinge cioè e non su quello che frena.

Vengo al dunque.

Agosto è un buon mese per fare SEO.

C'è il tempo di ragionare, chi si occupa dei contenuti generalmente è in ferie e i siti sono abbastanza fermi. Insomma, le interferenze sono al loro minimo possibile.

Agli inizi di agosto ho iniziato ad affrontare il sito www.multiplayer.it.

Si tratta di un portale di giochi online molto vasto, probabilmente il principale portale di giochi online italiano. Il traffico è elevato, solido, ma il cliente percepisce qualche difficoltà, nessun crollo ma un leggero calo, piuttosto costante. Il numero di pagine è molto alto. Il sito possiede molti terzi livelli dedicati ai singoli giochi, ha una robusta sezione cinema, molto aggiornata, e un canale di news in sofferenza: le news faticano a comparire in Google News. I contenuti sono unici, ottimi, ben scritti da redazioni specializzate e aggiornatissimi. Inoltre ha due aree con contenuti generati dagli utenti: un forum (i link nel forum hanno l'attributo rel=nofollow) e un terzo livello nel quale gli utenti potevano costruirsi un blog personale. I blog non hanno l'attributo rel=nofollow, sono però bloccati in pubblicazione da **molto tempo** ed è previsto il loro abbattimento definitivo agli inizi dell'anno prossimo, con la nuova versione del sito.

Perché Multiplayer aveva bisogno di un SEO? Perché è prevista una migrazione di CMS, e siccome il traffico da motori rappresenta una quota molto rilevante del traffico complessivo ecco che il cliente voleva fare le cose per bene, evitando i problemi che spesso capitano in occasioni simili.

Ad una prima analisi emergevano sostanzialmente due macro-problemi, subito individuati e oggetto di intervento:

- Nel sito veniva usata una forma di advertising interstiziale, simile a quella presente in Repubblica.it, ma senza redirectione automatica. In pratica dalla homepage tutti i link che portavano alle pagine interne di fatto dirigevano ad una pagina di pubblicità, ed era l'utente che doveva proseguire la navigazione cliccando un link che indirizzava all'articolo. La modalità di erogazione dell'advertising interstiziale rappresentava un ostacolo per gli spider, che potevano navigare il sito solo nei momenti in cui le campagne erano ferme, pochi giorni al mese. Tecnicamente venivano usati i cookie, e in assenza di cookie la pagina di pubblicità interstiziale veniva offerta stabilmente ad ogni articolo, invece che centellinata. Le difficoltà di Google News venivano subito imputate a questo problema.
- I blog erano stati **vandalizzati da link a siti spam** prima che fosse bloccata la possibilità di pubblicazione. Ma lo spam era rimasto, e Google lo vedeva. Apparentemente non c'erano ban sul terzo livello né altrove, e neppure penalizzazioni evidenti. C'era anzi stato un periodo molto buono nell'inverno precedente (con lo spam presente), con ottime rese sia per il traffico da motori che per utenti unici. Al punto che lo spam non era percepito da nessuno come

problema. A dirla tutta non era stato neppure visto. Emerge ad una semplice query *site:blog.dominio.it pharmacy | cialis | gambling*, ossia chiedendo a Google in quali pagine del sito fossero presenti le parole chiave *pharmacy | cialis | gambling* più altre estratte a campione dallo spam che quotidianamente affligge una normale casella di posta.. Un emulatore di spider lanciato sui primi blog evidenziati da queste query faceva emergere il carattere sistematico del vandalismo. Da sottolineare che in punta di logica era piuttosto complicato ascrivere il leggero calo di traffico degli ultimi mesi a un fattore presente anche durante il buon picco invernale: se lo spam è un problema, ed è presente da molto tempo, perché si manifesta solo ora? Logico, no? Apparentemente non fa una piega. Apparentemente.

Da segnalare che il settore giochi risente moltissimo delle “mode”: un gioco è di moda durante il suo lancio, quindi le ricerche si stabilizzano per un periodo per poi decrescere in modo naturale, a prescindere dai posizionamenti. L’arco di tempo è generalmente di pochi mesi, poi il pubblico che gioca è fidelizzato e non transita più per i motori. In queste condizioni ha poco senso fissare un pacchetto di parole chiave core-business con le quali misurare posizionamento e traffico, poiché l’interferenza del fattore “moda” (e la curva naturale del traffico) impedisce valutazioni sensate. Sotto un andamento di traffico complessivo piatto succedono a volte un sacco di cose.

Viene scelta un’altra strada, stare distanti dal [ticky-tacky](#) seo, irrobustire il sito per preparare la migrazione di CMS, renderlo quindi pienamente spiderizzabile e ripulirlo dallo spam. In altre parole: permettere al sito prima della migrazione di dispiegare pienamente il proprio potenziale, ora valutato compresso. Se non hai di fronte la curva naturale, effettiva e piena del traffico, non potrai capire se quello che fai è corretto o meno, se ha successo o meno. Guiderai insomma alla cieca e imputerai i risultati a questa o quella leva SEO in modo arbitrario.

Bene. Vi espongo i risultati degli interventi prospettati ed effettuati.

Il **primo punto** veniva risolto con una regola per gli user-agent che non accettano i cookie. Il default viene invertito: se non accetti i cookie non avrai mai la pubblicità interstiziale. Prima ce l’avevi sempre. Immediatamente riprende il flusso di news in Google News: da una dozzina a oltre 500 nel giro di pochi giorni. Questa questione viene risolta definitivamente nella prima decade di agosto, da allora il flusso di news in Google è costante e normale.

Il **secondo punto** prevedeva la soluzione canonica: l’abbattimento via Google Webmaster Tools dell’intero terzo livello contenitore dello spam, anticipando quanto era comunque previsto a gennaio. Quindi nell’ordine: sitemap.xml e file di controllo per aprire il pannello, robots.txt con disallow: / e rimozione in blocco del terzo livello dal database di Google. Il tutto avveniva il giorno **20 agosto**, a uffici chiusi, da un campeggio in Slovenia in riva a un lago. ☺

Ecco due grafici che illustrano meglio di ogni discorso le ripercussioni degli interventi.

Vengono usati due tool: il classico Google Analytics per il traffico da motori e Site Census di Nielsen per i visitatori unici.





Sottolineo che tutto avviene a bocce praticamente ferme, o quasi.

Nessuna attività significativa nel sito era in essere il 20 Agosto, nei giorni precedenti e in quelli successivi.

L'incremento di traffico è stato molto **netto e istantaneo**, tanto che quando al rientro dalle ferie ho contattato il cliente per notificargli quanto stava succedendo, questi mi ha detto che non c'era bisogno di nessun grafico per apprezzarlo: si erano riuniti d'urgenza dopo il 20 per negoziare un aumento di banda, visto che avevano raggiunto la saturazione.

Son soddisfazioni ☺

Sono anche indicazioni di un comportamento dell'algorithm molto diverso rispetto al 2006, quando le penalizzazioni erano evidenti e riguardavano il dominio in quanto tale.

Le penalizzazioni, i gravami da spam e il ticky-tacky SEO

Negli ultimi due anni le modalità con le quali Google ha erogato i cartellini (rossi e gialli, ban e penalizzazioni) sono mutate diverse volte, è imho c'è una logica stringente in questo.

Noi SEO cerchiamo sempre di prendere le misure al motore – per scopi più o meno nobili - ma il motore non vuole che gli prendiamo le misure. Ha un caratteraccio, si divincola, cambia comportamento.

Google ora non ti lascia più segni certi della penalizzazione, non c'è più un pattern secco individuabile (tranne che per il ban, ovvio). Ora ti sgonfia le gomme e ti lascia correre, ti dissangua lentamente e (non so se sia un'astuzia o un epifenomeno) ti lascia incaprettare da solo.

Se non sai di essere penalizzato tenderai a rafforzare il SEO delle pagine, scontrandoti facilmente con altre penalizzazioni (OOP, la sovraottimizzazione). Più ti dibatterai negli on-page e più rischierai di affondare nelle SERP.

Non più effetto sandbox, piuttosto **effetto quicksand**, per restare sulle sabbie (mobili).

La cosa ha anche altre implicazioni. Ad esempio se non sai quale nodo della matrice di link è certamente penalizzato non saprai più quale è certamente efficace. Sono le matrici di link (imho) il target principale di questa nuova forma di erogazione delle penalizzazioni.

Quando apparve l'evidenza della – 30 penalty (e poi della – 9xx) i SEO e webmaster più avveduti iniziarono a riconsiderare il problema dei link in uscita. La regola aurea era (ed è) che il motore ti considera responsabile in solido dei tuoi link in uscita, e quindi è cosa buona, giusta e fonte di salvezza stare molto attenti a chi segnali tramite un link.

Ma se in una matrice di link spam i nodi penalizzati sono agevolmente individuabili, la matrice potrà essere bonificata o aggiustata per prove ed errori, viceversa sarà molto più difficile e il format tenderà ad essere abbandonato perché inefficace, perché il rapporto costi/benefici sarà quasi sempre negativo.

Gli owner del singolo sito potranno sterilizzare per metodo col rel=nofollow, oppure controllare - a mano o tramite script - l'intero sito, o persino cambiare CMS, perché non puoi mettere continuamente patch e subire ugualmente attacchi SQL injection (vero PHPnuke?).

Ma l'owner di una matrice di link erosa da una penalizzazione lenta, silente e non individuabile cosa può fare, se non rivedere completamente l'approccio?

Quello che possiamo dire, qui e ora, è che la forma attuale delle penalizzazioni sembra definibile più come **gravame** che colpisce gruppi di parole chiave in base alla loro appetibilità, piuttosto che il sito in quanto tale. Come se Google ti mettesse uno zaino pesante sulle spalle, ma ti lasciasse comunque correre. Come se ti sgonfiasse le ruote, ma ti lasciasse gareggiare.

“Lo spam sta meglio in SERP, ma fiacco (oltre la terza pagina), piuttosto che dentro e fuori, dando continue chance allo spammer di migliorarsi e fare fine tuning”: così sembra ragionare oggi il principale motore di ricerca.

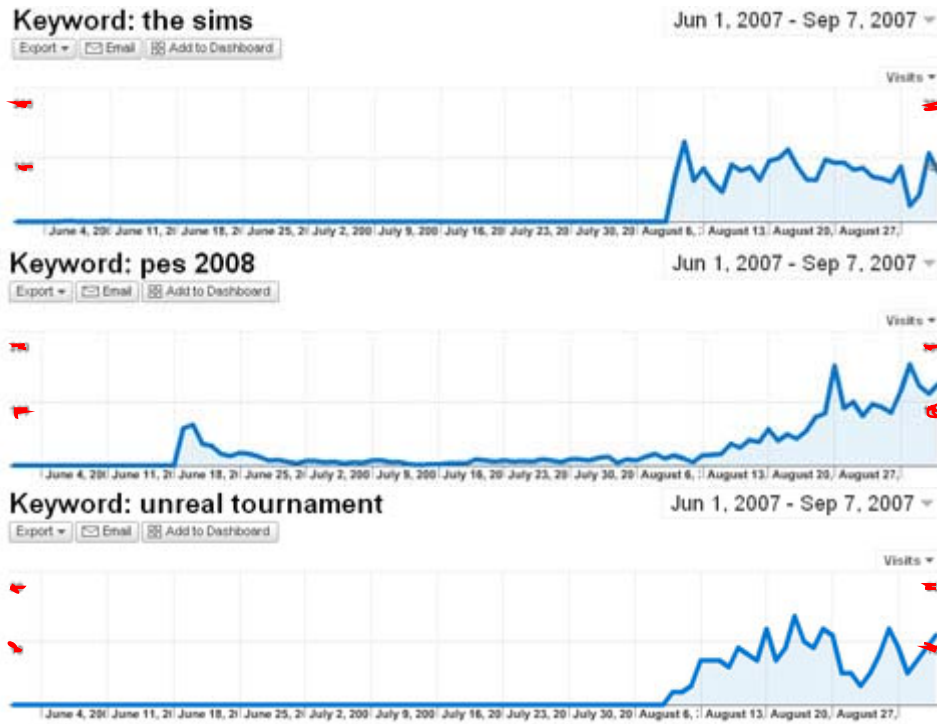
Conseguenze pratiche per i SEO? Molte. Ognuno tragga le sue. Vi dico quelle che ho tratto per me.

- Le check list di controlli costruite finora per verificare le penalizzazioni si possono buttare, le nuove penalizzazioni sono asintomatiche, almeno per quello che riguarda il nome dominio.
- Chiunque abbia (e persino abbia avuto) contenuti generati da utenti *_deve_* affrontare il problema dello spam. Il ventaglio delle possibilità è vasto: dalla moderazione attenta e continua, alle black list dei termini che indicano attività di spamming, fino alla sterilizzazione metodica dei link in uscita con `rel=nofollow`. Si può scegliere, ma non ha alcuna importanza che lo spam sia vecchio o fresco, va estirpato a prescindere.
- L'anamnesi di un sito assume un ruolo centrale: molto prima di applicare le liturgie SEO classiche va fatta un'analisi sullo storico del sito e una valutazione sul suo potenziale. D'ora in avanti mi prendo almeno una mezza giornata per chiacchierare rilassato sulle variazioni che il sito ha avuto dalla nascita in avanti. Il cliente va lasciato parlare per ore se possibile, nessuna check list preordinata di domande perché le check list danno una errata sensazione di sicurezza, che tutto sia stato valutato una volta che è stato fatto il segno di spunta sull'ultima voce.
- Il ticky-tacky SEO, ovvero il SEO di routine, sempre uguale a se stesso, fatto per interventi standard principalmente sugli on-page ha fatto il suo tempo. Si può fare SEO senza neppure toccare le leve SEO classiche, almeno in prima battuta. L'importanza dei **fattori wide-site** emerge con prepotenza da casi come questo, così come la consapevolezza che si può lavorare anche per sottrazione e non solo per addizione.
- Se le penalizzazioni sono una malattia subdola e silenziosa, questo non significa che tu non possa scoprirla, ma la certezza ce l'avrai solo quando ne sei uscito. Vi sottopongo il rendimento nel tempo di alcune parole chiave di Multiplayer, quelle che vengono in mente a me, che ho fatto l'ultima partita a Duke Nukem il secolo scorso e sui giochi sono una pippa totale.

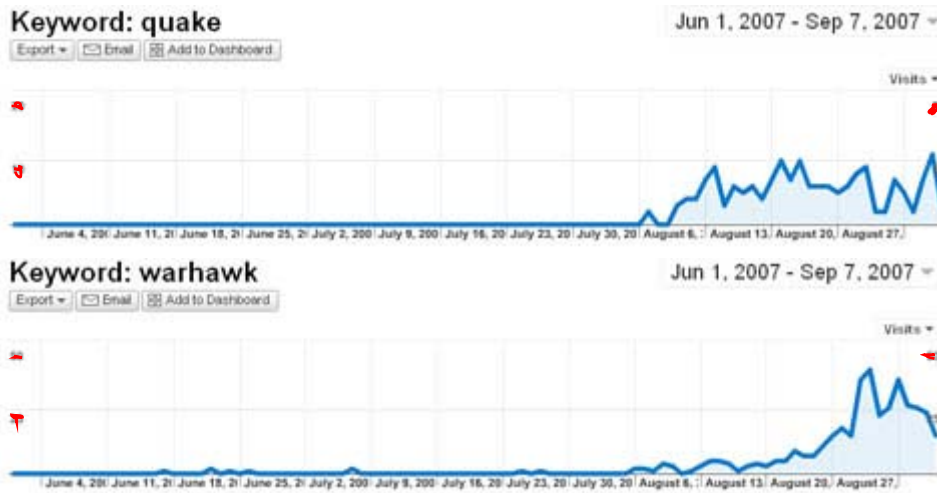
Una console:



Tre giochi famosi:

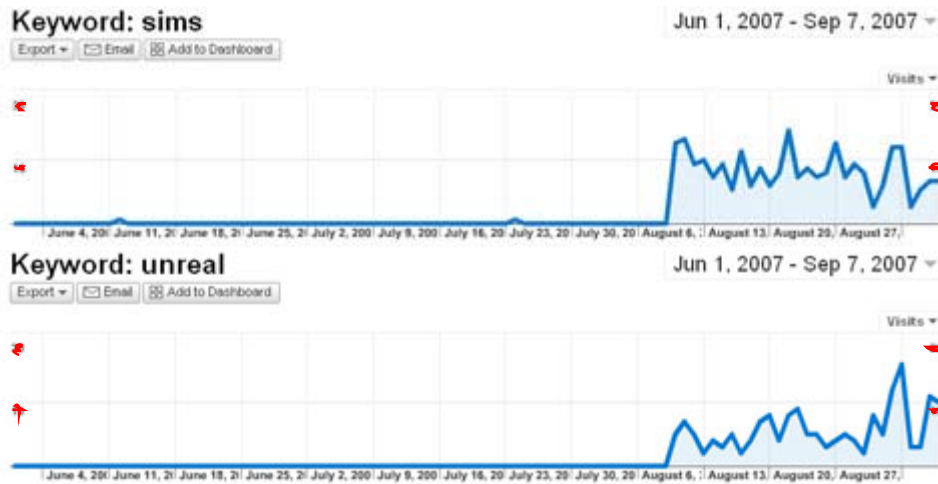


Due giochi famosi, a parola singola:



ì
Tre parole chiave singole, contenute in nomi di gioco ma polisemantiche:





Andamenti simili sono stati riscontrati in tutta la coda lunga, che inizia a muoversi dopo il 5 agosto (con l'inversione della regola dell'advertising interstiziale) ma **decolla in modo improvviso** - e nel suo insieme - il giorno 20.

Quello che si dice un risveglio in piena regola, che vale in termini di traffico complessivo attorno a un +30% di visitatori unici.

Per chi volesse approfondire i cambiamenti delle forme di penalizzazione segnalo due thread:

Vecchia forma di penalizzazione secca sul dominio: [il caso Agendaonline](#) ott-dic 2006

Nuova forma di penalizzazione sulle kw: la case history di Giorgio Tave su [Libri-online](#) febb-giu 2007

Si ringrazia Multiplayer.it per aver concesso di trattare il caso pubblicamente, dimostrando così di possedere una webness non comune, che certamente sarà premiata dai motori di ricerca ☺

Si ringrazia lo spirito di Malvina Reynolds per il concetto di ticky-tacky

*Little boxes on the hillside,
Little boxes made of ticky-tacky,
Little boxes, little boxes,
Little boxes, all the same.
There's a green one and a pink one
And a blue one and a yellow one
And they're all made out of ticky-tacky
And they all look just the same.*