

[0045] *Content Update/Change*

[0045] Aggiornamenti e cambiamenti dei contenuti

[0046] According to an implementation consistent with the principles of the invention, information relating to a manner in which a document's content changes over time may be used to generate (or alter) a score associated with that document. For example, a document whose content is edited often may be scored differently than a document whose content remains static over time. Also, a document having a relatively large amount of its content updated over time might be scored differently than a document having a relatively small amount of its content updated over time.

[0046] In una applicazione coerente con i principi dell'invenzione, informazioni relative al modo in cui **i contenuti di un documento cambiano** nel tempo possono essere usate per generare (o alterare) un punteggio associato con il documento stesso. Per esempio, ad un documento i cui contenuti vengono editati spesso può essere assegnato un punteggio differente da quello assegnato ad un documento i cui contenuti non vengono aggiornati e rimangono statici nel tempo. Così come un documento che nel tempo aggiorni gran parte del suo contenuto potrebbe ricevere un punteggio diverso rispetto ad un documento che nel tempo ne aggiorni solo una piccola parte.

[0047] In one implementation, search engine 125 may generate a content update score (U) as follows:

[0047] In un'implementazione il motore di ricerca può generare un punteggio relativo all'aggiornamento dei contenuti come segue:

[0048] $U=f(UF, UA)$,

[0048] $U=f(UF, UA)$, dove:

U= punteggio assegnato al documento in funzione di:

UF=frequenza di aggiornamento dei contenuti

UA=quantità dei contenuti aggiornati

[0049] where f may refer to a function, such as a sum or weighted sum, UF may refer to an update frequency score that represents how often a document (or page) is updated, and UA may refer to an update amount score that represents how much the document (or page) has changed over time. UF may be determined in a number of ways, including as an average time between updates, the number of updates in a given time period, etc.

[0049] dove f può essere una funzione, come una somma o una somma pesata, UF un punteggio

A cura di:
Stefano Becheroni di [Posizionamento Web](#)
Francesco Iamurri di [Motori e Ricerca](#)

legato alla **frequenza degli aggiornamenti**, che rappresenta quanto spesso un documento (o una pagina) viene aggiornato, e UA può essere un punteggio legato alla **quantità degli aggiornamenti**, che rappresenta quanto il documento (o la pagina) è cambiato nel tempo. UF può essere determinato in vari modi, per esempio da una media del tempo che passa tra gli aggiornamenti, dal numero degli aggiornamenti effettuati in un dato periodo di tempo, etc.

[0050] UA may also be determined as a function of one or more factors, such as the number of "new" or unique pages associated with a document over a period of time. Another factor might include the ratio of the number of new or unique pages associated with a document over a period of time versus the total number of pages associated with that document. Yet another factor may include the amount that the document is updated over one or more periods of time (e.g., n % of a document's visible content may change over a period t (e.g., last m months)), which might be an average value. A further factor might include the amount that the document (or page) has changed in one or more periods of time (e.g., within the last x days).

[0050] UA può anche essere determinato come funzione di uno o più fattori, per esempio il numero di "nuove pagine" o di pagine uniche associate con un documento in un dato periodo. Un altro fattore potrebbe includere il rapporto fra le "nuove pagine" o le pagine uniche associate con il documento in un periodo di tempo e il numero di pagine totale associate con lo stesso. Un altro fattore ancora può includere la quantità del documento che viene aggiornata in uno o più periodi di tempo (per esempio, n% del contenuto visibile di un documento può cambiare in un determinato tempo t (gli ultimi m mesi)), che potrebbe essere considerata un valore medio. Un ulteriore fattore potrebbe includere la quantità del documento (o della pagina) che è cambiata in uno o più periodi di tempo (entro gli ultimi x giorni).

[0051] According to one exemplary implementation, UA may be determined as a function of differently weighted portions of document content. For instance, content deemed to be unimportant if updated/changed, such as Javascript, comments, advertisements, navigational elements, boilerplate material, or date/time tags, may be given relatively little weight or even ignored altogether when determining UA. On the other hand, content deemed to be important if updated/changed (e.g., more often, more recently, more extensively, etc.), such as the title or anchor text associated with the forward links, could be given more weight than changes to other content when determining UA.

[0051] In una tipica implementazione, UA può essere determinata come una funzione di diverse porzioni di contenuto di un documento pesate differentemente. Per esempio, i cambiamenti di contenuti considerati non importanti, come i javascript, i commenti, la pubblicità, gli elementi di navigazione o i tag data/giorno, possono avere un peso relativamente piccolo o essere ignorati nella determinazione di UA. D'altra parte, i contenuti ritenuti importanti se cambiati/aggiornati (più spesso, più recentemente, più estensivamente, etc), come il titolo o il testo dei link uscenti, possono avere un peso maggiore dei cambiamenti degli altri contenuti nel determinare UA.

[0052] UF and UA may be used in other ways to influence the score assigned to a document. For example, the rate of change in a current time period can be compared to the rate of change in another (e.g., previous) time period to determine whether there is an acceleration or deceleration trend. Documents for which there is an increase in the rate of change might be scored higher than those documents for which there is a steady rate of change, even if that rate of change is relatively high. The amount of change may also be a factor in this scoring. For example, documents for which there is an increase in the rate of change when that amount of change is greater than some threshold might be scored higher than those documents for which there is a steady rate of change or an amount of change is less than the threshold.

A cura di:
Stefano Becheroni di [Posizionamento Web](#)
Francesco Iamurri di [Motori e Ricerca](#)

"Tutti i testi e i contenuti originali del documento sono di proprietà dei rispettivi autori. Il presente documento è pubblicato secondo la [licenza di utilizzo di Creative Commons](#), [qui riassunta](#), per facilitarne la diffusione. Stefano Becheroni e Francesco Iamurri non si assumono alcuna responsabilità nel caso di eventuali errori contenuti negli articoli o di errori in cui si fosse incorso nella loro riproduzione sul sito."

[0052] UF and UA possono essere usati in altri modi per influenzare il punteggio assegnato a un documento. Per esempio: il ritmo dei cambiamenti riferito ad un periodo di tempo attuale può essere paragonato col ritmo dei cambiamenti in un altro (precedente) periodo per determinare se ci sono tendenze di accelerazione o decelerazione dello stesso. Documenti per i quali vengano rilevati incrementi nel ritmo dei cambiamenti potrebbero avere un punteggio maggiore di quei documenti il cui ritmo dei cambiamenti rimane costante, anche se relativamente alto. Anche l'entità di cambiamenti può rappresentare un fattore nell'assegnazione di questo punteggio. Per esempio, documenti per i quali c'è un incremento nel tasso di cambiamento e nei quali l'entità dei cambiamenti supera una certa soglia potrebbe ricevere un punteggio maggiore di quei documenti per i quali il tasso di cambiamento rimane costante o nei quali l'entità dei cambiamenti rimanga inferiore alla suddetta soglia.

[0053] In some situations, data storage resources may be insufficient to store the documents when monitoring the documents for content changes. In this case, search engine 125 may store representations of the documents and monitor these representations for changes. For example, search engine 125 may store "signatures" of documents instead of the (entire) documents themselves to detect changes to document content. In this case, search engine 125 may store a term vector for a document (or page) and monitor it for relatively large changes. According to another implementation, search engine 125 may store and monitor a relatively small portion (e.g., a few terms) of the documents that are determined to be important or the most frequently occurring (excluding "stop words").

[0053] In alcune situazioni, le risorse per la memorizzazione dei documenti possono essere insufficienti per archiviare i documenti allo scopo di monitorarne i cambiamenti dei contenuti. In questo caso il motore di ricerca può archiviare delle "**rappresentazioni**" dei documenti e monitorare queste ultime. Per esempio il motore di ricerca può immagazzinare delle "signature" dei documenti invece degli interi documenti per scoprire cambiamenti nei contenuti. In questo caso il motore di ricerca archivia un "vettore di termini" per un documento (o una pagina) e lo controlla per rilevare cambiamenti relativamente grandi nel documento. In un'altra implementazione, il motore di ricerca può immagazzinare e monitorare una parte relativamente piccola (per esempio pochi termini) del documento che è ritenuta importante o più frequentemente ricorrente (escludendo le "stop words").

[0054] According to yet another implementation, search engine 125 may store a summary or other representation of a document and monitor this information for changes. According to a further implementation, search engine 125 may generate a similarity hash (which may be used to detect near-duplication of a document) for the document and monitor it for changes. A change in a similarity hash may be considered to indicate a relatively large change in its associated document. In other implementations, yet other techniques may be used to monitor documents for changes. In situations where adequate data storage resources exist, the full documents may be stored and used to determine changes rather than some representation of the documents.

[0054] In un'implementazione, il motore di ricerca può immagazzinare un **sommario o un'altra rappresentazione** di un documento e monitorare questa informazione per rilevare cambiamenti. In un'ulteriore implementazione, il motore di ricerca può generare un **codice di somiglianza** (che può essere usato per determinare duplicazioni e simili di un documento) per il documento e monitorare quello per rilevare i cambiamenti. Un cambiamento in un codice di somiglianza può essere considerato indicativo di un cambiamento rilevante del documento associato. In altre implementazioni, possono essere usate anche ulteriori tecniche per monitorare i documenti allo scopo di determinarne i cambiamenti. In situazioni dove esistono adeguate risorse per la memorizzazione dei dati, invece di una rappresentazione può essere immagazzinato e usato l'intero documento, per monitorare i cambiamenti dello stesso.

A cura di:
Stefano Becheroni di [Posizionamento Web](#)
Francesco Iamurri di [Motori e Ricerca](#)

[0055] For some queries, documents with content that has not recently changed may be more favorable than documents with content that has recently changed. As a result, it may be beneficial to adjust the score of a document based on the difference from the average date-of-change of the result set. In other words, search engine 125 may determine a date when the content of each of the documents in a result set last changed, determine the average date of change for the documents, and modify the scores of the documents (either positively or negatively) based on a difference between the documents' date-of-change and the average date-of-change.

[0055] Per alcune ricerche possono essere favoriti documenti che non sono stati recentemente aggiornati rispetto ad altri aggiornati più di recente. Quindi può essere conveniente aggiustare il punteggio di un documento basandosi sulla differenza dalla media delle date dei cambiamenti dei documenti in un insieme di risultati. In altre parole, il motore di ricerca può determinare la data in cui **il contenuto di ognuno dei documenti in un insieme di risultati è stato cambiato**, e modificare il punteggio dei documenti (sia positivamente che negativamente) basandosi sulla differenza tra la data di cambiamento di ciascun documento e la data di cambiamento media.

[0056] In summary, search engine 125 may generate (or alter) a score associated with a document based, at least in part, on information relating to a manner in which the document's content changes over time. For very large documents that include content belonging to multiple individuals or organizations, the score may correspond to each of the sub-documents (i.e., that content belonging to or updated by a single individual or organization).

[0056] In sintesi il motore di ricerca può generare (o alterare) un punteggio associato con un documento basandosi, almeno in parte, su informazioni relative al modo in cui il contenuto del documento cambia nel tempo. Per documenti molto grandi che includono contenuti appartenenti a più individui o organizzazioni, il punteggio può essere assegnato a ciascun sub-documento (che per esempio appartiene a un singolo individuo o organizzazione o che è aggiornato da un singolo individuo o organizzazione).

A cura di:
Stefano Becheroni di [Posizionamento Web](#)
Francesco Iamurri di [Motori e Ricerca](#)