

Dokumentation Volltextsuche Implisense



29. März 2017

In dieser Kurzdokumentation sollen die grundlegenden Funktionen der Volltextsuche bei Implisense anhand von Beispielen beschrieben werden. Um das Verhalten der Suchmaschine zu demonstrieren, verwenden wir die Suchterme `lead` und `generation`.

1 Anfrage: `lead`

Im einfachsten Fall gibt man einen einzelnen Term in die Suche ein. Für das Suchergebnis müssen wir die Filterung und die Sortierung unterscheiden. In diesem einfachsten Falle müssen alle Firmen (deren Stammdaten oder Inhalte der Webseiten), die das Suchergebnis ausmachen, den Term enthalten. Alle anderen Firmen werden ausgefiltert. Groß- und Kleinschreibung der Terme wird ignoriert. Die Sortierung des Suchergebnisses, also die Reihenfolge, in der die Firmen präsentiert werden, hängt von der Relevanz der Firma für die Anfrage ab. Diese Relevanz ergibt sich daraus, wie oft der Suchterm bei den Inhalten der Firma auftaucht, wobei diese Anzahl hinsichtlich der Größe der Inhalte normalisiert wird. Stand März 2017 liefert die Anfrage 20348 Firmen als Suchergebnis.

2 Anfrage: `lead generation`

In der einfachsten Form einer Anfrage mit zwei Termen werden beide Terme durch ein Leerzeichen getrennt eingegeben. Hinsichtlich der Filterung bedeutet dies bei Implisense ein logisches ODER. D. h. das Suchergebnis enthält alle Firmen, deren Inhalte mindestens einen der beiden Terme enthalten. Das gleiche Ergebnis kann auch über die explizite Eingabe des OR-Operators `|` erzielt werden, also über die Anfrage `lead | generation`. Hinsichtlich der Sortierung wird wieder die Relevanz berücksichtigt. Firmen, deren Inhalte beide Terme enthalten, haben üblicherweise Priorität über Firmen, die nur einen der beiden Terme enthalten. Firmen, deren Inhalte beide Terme direkt hintereinander in der angegebenen Reihenfolge enthalten, haben eine noch höhere Priorität. Auch die Anzahl des Auftretens der Terme spielt wie oben wieder eine Rolle für die Relevanz. Stand März 2017 liefert die Anfrage 95611 Firmen als Suchergebnis.

Vorsicht! Die Volltextsuche von Implisense unterstützt nicht die bekannten Suchoperatoren AND, OR und NOT. Die Anfrage `lead OR generation` liefert also

sehr viele Ergebnisse, da jede Firma, deren Inhalte einen der drei Terme dieser Anfrage enthalten, Teil des Suchergebnisses ist.

3 Anfrage: lead + generation / +lead +generation

Wenn das Auftreten bestimmter Terme erzwungen werden soll, so sollten diese mit einem + verbunden oder direkt eingeleitet werden. Vor dem ersten Term kann das + weggelassen werden, so dass die weiteren verbindenden + einem logischen UND entsprechen (erste Form). Alternativ können alle Terme direkt mit dem + eingeleitet werden (zweite Form). Das Suchergebnis enthält dann alle Firmen, deren Inhalte beide Terme enthalten. Wiederum werden die Firmen mit der höchsten Relevanz zuerst im Suchergebnis geliefert. Stand März 2017 liefert die Anfrage 7672 Firmen als Suchergebnis.

Vorsicht! Die Volltextsuche von Implisense unterstützt nicht die bekannten Suchoperatoren AND, OR und NOT. Die Anfrage `lead AND generation` kann also nicht verwendet werden, um das identische Suchergebnis zu erzielen.

4 Anfrage: ‘lead generation’

Mit Hilfe der Phrasensuche kann das direkte Aufeinanderfolgen von Termen erzwungen werden. Das Suchergebnis enthält also nur Firmen, deren Inhalte nicht nur beide Terme, sondern beide Terme direkt hintereinander enthalten. Stand März 2017 liefert die Anfrage 370 Firmen als Suchergebnis (im Vergleich zu 7672 für die normale UND-Suche).

5 Anfrage: -lead

Bestimmte Terme können auch ausgeschlossen werden. Im Gegensatz zum + wird das - am Anfang der Anfrage nicht ignoriert. Man kann also Terme ausschließen, ohne andere Terme, die auftreten müssen, anzugeben. Stand März 2017 liefert die Anfrage 1749830 Firmen als Suchergebnis.

6 Anfrage: lead + -generation

Erzwungenes Auftreten und Nicht-Auftreten von Termen kann kombiniert werden. Diese Anfrage liefert Firmen, deren Inhalte den ersten, nicht aber den zweiten Term enthalten. Stand März 2017 liefert die Anfrage 12676 Firmen als Suchergebnis.

Vorsicht! Das +-Zeichen in der Anfrage ist wichtig, denn es zeigt das logische UND an. Die Anfrage `lead -generation` impliziert ein logisches ODER zwischen den Termen, d. h. es werden alle Firmen geliefert, deren Inhalte den ersten Term enthalten oder den zweiten Term nicht enthalten.

7 Anfrage: lead + generat*

Eine Vervollständigung von Termen mit Wildcards ist ebenfalls möglich. Wenn also neben `generation` auch noch `generator` als zulässiger Term gilt, kann die Anfrage wie oben formuliert werden. Wildcards funktionieren nicht in einer Phrasensuche. Stand März 2017 liefert die Anfrage 12598 Firmen als Suchergebnis (im Vergleich zu 7672 für die Suche ohne Wildcards).

8 Weiteres / Fortgeschrittenes

Eine weitere Form, mit nicht genau spezifizierten Termen umzugehen, ist die Suche über Editierdistanz. Die zulässige Editierdistanz kann über eine Tilde `~` angegeben werden. Formuliert man z. B. die Anfrage `maschine~1`, so deckt diese sowohl den deutschen Plural `maschinen` als auch das englische `machine` ab, da jeweils nur eine Editierung (Hinzufügen des `n` am Ende, Weglassen des `s`) nötig ist, um den Ausgangsterm in den jeweils anderen umzuwandeln.

Verwendet man mehr als drei Terme in der Anfrage, so können runde Klammern verwendet werden, um die verschiedenen Operatoren nach Priorität zu gruppieren (z. B. `lead + (generation | scoring)`).

Weitere Informationen zur Elasticsearch Simple Query Syntax, die von uns verwendet wird, können auch der englischsprachigen Originaldokumentation entnommen werden.