

Semantic Search

Im Gegensatz zu konventioneller Volltextsuche verwendet Ontonyms Semantic Search Hintergrundwissen, um passende Dokumente zu ermitteln und Dokumente herauszufiltern, die herkömmliche Suchmaschinen fälschlicherweise als passend betrachten würden.



Hierbei hängt die Passgenauigkeit der Dokumente von der Ähnlichkeit der verwendeten Begriffe im Dokument und der Suchanfrage ab.

Je nach Hintergrundwissen werden so einerseits Dokumente findbar, die z.B. anstelle des Suchbegriffs „Abschreibungstabelle“ den synonymen Begriff „AfA-Tabelle“ enthalten oder die anstelle des Suchbegriffs „Gehaltszulage“ speziellere Arten von Zulagen beschreiben, wie z.B. „Kinderzuschlag“ oder „Weihnachtsgeld“. Andererseits werden Fehltreffer wie z.B. die Anzeige der „Assistentin des Geschäftsführers“ bei der Suche nach „Geschäftsführer“ oder des Dokuments mit Titel „Serviceorientierte Architektur für Kunde XYZ“ bei der Suche nach „CEO“ vermieden.

Die Ergebnisse der semantischen Suche sind im Wesentlichen vom Umfang und der Qualität des verwendeten Hintergrundwissens abhängig. Dieses Hintergrundwissen sollte den Sprachgebrauch der Autoren und der Suchenden abdecken. Da beide Gruppen jedoch bei der Wahl ihrer Suchbegriffe sehr kreativ sind und mitunter neue Worte kreieren und verwenden, die noch nicht im Hintergrundwissen erfasst wurden, greift Ontonyms semantische Suche im Notfall auch mal auf eine konventionelle Volltextsuche zurück, um das Informationsbedürfnis des Benutzers zu befriedigen.

Die Modellierung und Pflege des Hintergrundwissens erfordert den Umgang mit komplexen, für Entwickler ausgelegten speziellen Modellierungswerkzeugen (d.h. Ontologieeditoren) und erfordert einige Erfahrung im Bereich der Wis-

sensmodellierung. Um hierbei nicht von eigenen Ressourcen abhängig zu sein, kann Sie Ontonym selbstverständlich unterstützen. Wir verfügen nicht nur über die nötige Modellierungserfahrung, sondern auch über Analysewerkzeuge zur effizienten Gestaltung des Modellierungsprozesses.

Eigenschaften

Java-basiert, Nutzung von JBoss Application Server (Verwendung von anderen Application Servern möglich), relationale DB: Standardmäßig MySQL oder PostgreSQL (oder andere RDBS mit JDBC-Unterstützung)

○ Automatische Ontologie-basierte Verschlagwortung als Basis für die semantische Suche.

○ Berücksichtigung von Synonymen und Abkürzungen, Ober- und Unterbegriffsbeziehungen bei Suchanfragen.

○ Automatische Rechtschreibfehlerkorrektur von einfachen Tippfehlern anhand des Hintergrundwissens, wie z.B. Auslassung, Einfügung, Vertauskung, Buchstabendreher.

① Semantische Anfragevervollständigung. Während herkömmliche Suchmaschinen – wie auch Webbrowser – lediglich eine syntaktische Anfragevervollständigung ermöglichen, werden Benutzern von Ontonyms semantischer Suche inhaltliche Anfragevervollständigungen auf der Basis des verwendeten Hintergrundwissens bei der Eingabe zur Auswahl vorgeschlagen.

The screenshot shows the yapadu search engine interface. At the top, there is a search bar with the text 'Lehrer' entered. A dropdown menu is open, showing suggestions: 'Lehrer', 'Lehrer', 'Berufsschullehrer', 'Grund- u. Hauptschullehrer', 'Gymnasiallehrer', 'Realschullehrer', and 'Sportlehrer'. Below the search bar, it says '14 Stellen gefunden, 10161 insgesamt'. The main content area shows search results for 'Ähnliche Konzepte: Berufsschullehrer, Grund- u. Hauptschullehrer, Gymnasiallehrer, Realschullehrer, Sportlehrer, Grundschullehrer, Handelslehrer, Hauptschullehrer, Studienrat, Diplom-Handelslehrer, Oberstudienrat'. There are three search results listed, each with a title, a brief description, and a relevance percentage.

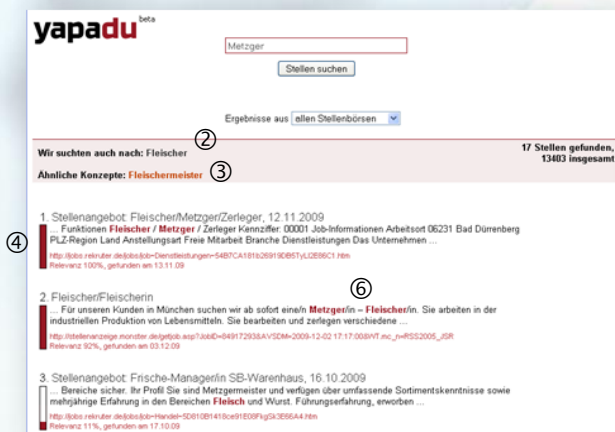
○ Intelligente Phrasenerkennung in der Anfrage. Wollen Benutzer bei konventionellen Volltextsuchmaschinen nach zusammenhängenden Begriffen in einer bestimmten Reihenfolge (diese werden oft als Phrase bezeichnet) suchen, stehen ihnen in der Regel Anführungszeichen zur Verfügung, um dies in der Anfrage zu kennzeichnen, z.B. „Marktstudie „Semantic Web““. An Hand der Ontologie analysiert Ontonyms semantische Suche Anfragen auf das implizite Vorhandensein zusammenhängender Begriffe und gewichtet diese stärker als eine einfache Und-Verknüpfung der Begriffe. So werden beispielsweise bei der Anfrage ‚Marktstudie Semantic Web‘ Dokumente, in denen „Semantic Web“ zusammenhängend auftritt, höher bewertet als Dokumente, in denen beide Begriffe nicht zusammenhängend auftreten.

- ④ Graphische Visualisierung der Relevanzbewertung.
- ⑤ Relevanzsortierte Ergebnisse.
- ⑥ Hervorhebung der gefundenen Suchbegriffe im Textauszug. Synonyme und ähnliche Unterbegriffe werden ebenfalls hervorgehoben.

Mögliche Anpassungen

Neben diesen Eigenschaften kann Ontonyms semantische Suche bei Bedarf flexibel erweitert werden, z.B. um:

- Integration anderer Volltextsuchmaschinen als „fall back“
- Sortierung nach Datum oder anderen Kriterien
- Rechtschreibfehlererkennung und Korrekturvorschläge bei mehreren Schreibfehlern
- Identifikation und Hervorhebung der signifikant relevantesten Treffer
- Boolesche Anfrage-logik
- Anpassung der Ergebnisdarstellung
- Hervorgehobene Anzeige von „featured links“, d.h. Verweisen auf wichtige Dokumente zu den gefundenen Suchbegriffen



○ Lucene-basierte Volltextsuche als „fall back“. Für den Fall, dass die semantische Suche einmal kein Hintergrundwissen zu einem Suchbegriff kennt.

○ Gewichtung von Begriffen. Strukturelle Informationen werden genutzt, um Vorkommen der Suchbegriffe in unterschiedlichen Teilen des Dokuments unterschiedlich zu gewichten, z.B. im Titel und im Dokumenttext. Die Gewichtung erfolgt in Abhängigkeit von der Vorkommenshäufigkeit. Weitere Anpassungen der Gewichtungen sind möglich.

② Anzeige der Synonyme, die bei der Suche mit betrachtet wurden.

③ Anzeige der Unterbegriffe, die bei der Suche berücksichtigt wurden.

Kontakt

Ontonym – Gesellschaft für
semantische Webanwendungen mbH
Bismarckstr. 18
14109 Berlin
Tel: 030-80403246
kontakt@ontonym.de
http://www.ontonym.de