

Special Innovation

Suchspiel mit Grips

Im digitalen Zeitalter wächst der verfügbare Content teilweise exponentiell. Damit konnten aber die bisherigen Suchstrategien nicht mehr Schritt halten. Neue Wege geht hier seit einiger Zeit erfolgreich die Austria Presse Agentur.

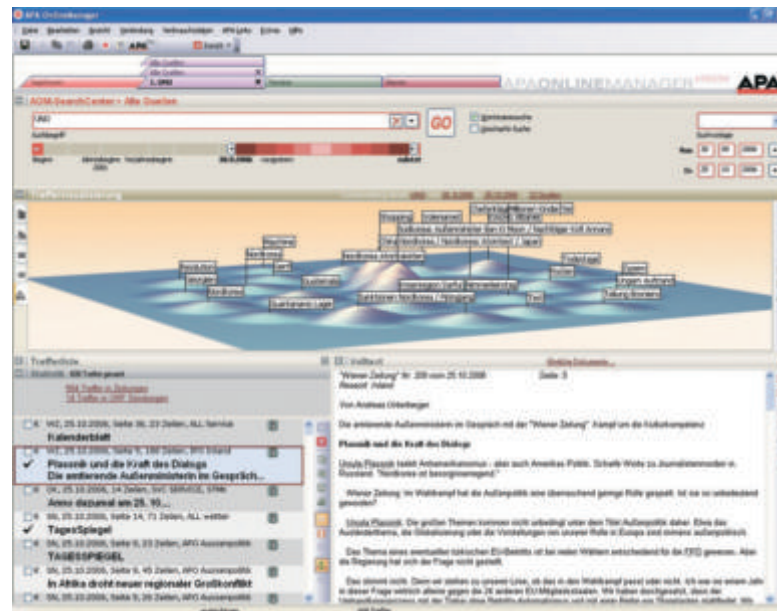
Ernst Brandstetter

König Content erfährt einen schleichenden Tod, versteckt in den unendlichen Weiten des Informationsuniversums. Doch die Rettung naht: Innovative Archiv- und Datenbanklösungen bieten dem User individuelle Wege zum Suchergebnis und bunte Bilder statt langweiliger Trefferlisten.

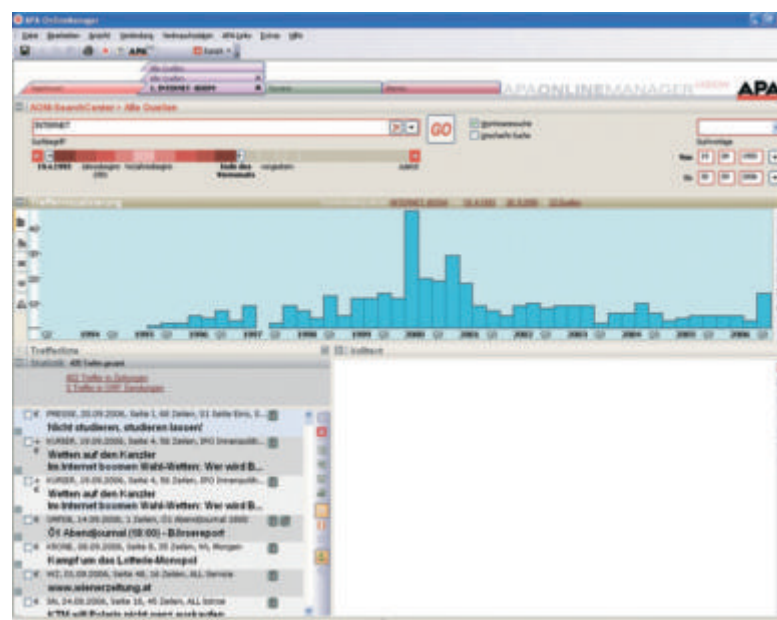
Hohe Anforderungen

Auf der Suche nach relevanter Information in den undurchdringlichen Strukturen großer Archive bleibt der Suchende oft ziemlich allein. Suchmaschinen liefern Ergebnisliste um Ergebnisliste – allein das Gesuchte bleibt verborgen und offenbart sich nur in Glücksfällen als die berühmte Stecknadel im Heuhaufen. Tragisch für den Gelegenheitssucher – untragbar für professionelle Anwender wie beispielsweise Journalisten, deren Story nicht

übermorgen, sondern zu Redaktionsschluss in zwei Stunden fertig sein muss. Medien-Archive sind Schatzgruben. Sie beinhalten oft Mio. von Text- und Bild-Dokumenten, nicht selten auch Audio-Files, und wachsen jeden Tag um Tausende neue Informationen. Herkömmliche Datenbanksysteme stoßen bei solchen Zahlen schnell an ihre Grenzen und liefern oftmals unbefriedigende Suchergebnisse. Vor diesem Problem stand auch die Apa – Austria Presse Agentur im Jahr 2003. Das bis dato eingesetzte Datenbanksystem hätte den zukünftigen Anforderungen des Agenturbetriebs nicht mehr gerecht werden können. „Das System soll für große Datenmengen ausgelegt sein, kurze Antwortzeiten ermöglichen und – besonders wichtig – dem Suchenden eine Vielzahl an Funktionen und Erleichterungen für eine effiziente Suche bieten“, erklärt Apa-IT Geschäftsführer Rudolf Horvath. „Allerdings stellte sich schnell heraus, dass auch neuere auf dem Markt erhältliche Datenbanken an den hohen Ansprüchen eines Rund-um-die-Uhr-Medienbetriebes scheitern mussten. Deshalb haben wir uns aufgrund unserer langjährigen Erfahrung im Bereich der Volltextdatenbanken entschlossen, ein eigenes System ganz nach unseren Wünschen zu entwickeln.“ Vor diesem Hintergrund konzipierte die Innovationsabteilung von Apa-IT in knapp 18 Monaten Entwicklungszeit das Datenbanksystem Power Search.



Themen-Clustering und Informationsvisualisierung im Apa-Online-Manager (AOM) mit Apa-IT-Power-Search. Foto: APA-IT



Anhand der Zeitreihen-Darstellung kann der Benutzer Thementrends über einen längeren Zeitraum verfolgen. Foto: APA-IT

Starkes Interesse

Power Search hat sich im Agenturbetrieb als Datenbanksystem des Apa-Online-Managers (AOM) bestens bewährt. „Innovationen, die funktionieren und begeistern, bleiben meist nicht unbemerkt. Deshalb haben auch bald nach der Einführung des Systems deutsche Medienbetriebe ihr Interesse an der Technologie der Apa-IT Media Archives angemeldet“, freut sich Geschäftsführer Rudolf Horvath. Die Deutsche Presse-Agentur (DPA) mit Sitz in Hamburg beschloss etwa, ihr Medienarchiv zur Apa auszulagern und für die Recherchen ihrer Journalisten zu optimieren. 14 Mio. Dokumente wurden in die Apa-IT-Power-Search-Datenbank übernommen und stehen nun 1.000 DPA-Usern zur Verfügung. Apa-IT erstellte für diese Lösung auch die Recherche-Oberfläche der DPA. Auf Basis des AOM wurde ein Web-Interface im Corporate Design der DPA entwickelt, über das neben den Recherchen auch das Accounting für externe zahlende Kunden der DPA abgewickelt werden kann.

Optimale Recherche

Anfang 2006 entschieden sich auch der Süddeutsche Verlag sowie Gruner + Jahr, ihre Medienarchive auf Power-Search-Technologie umzustellen und in die Obhut von Apa-IT zu geben. 18 Mio. Dokumente der beiden Unternehmen werden künftig bei Apa-IT gespeichert; auch hier bieten individuell angepasste Web-Oberflächen mit kundenspezifischen Spezialfunktionen optimale Recherchebedingungen. Der Boom innovativer Archiv- und Datenbanklösungen hat gerade erst begonnen – österreichische Unternehmen wie Apa-IT sind ganz vorne mit dabei. *bra*

Steckbrief



Rudolf Horvath ist Geschäftsführer von Apa-IT.
Foto: APA

Power Search im Info-Cluster

Verfügbares Wissen wird in Hügeln, Balken und Landkarten für Nutzer sichtbar gemacht.

Power Search kann in Sekundenbruchteilen mehr als 100 Mio. Dokumente (Text, Bilder, Grafiken und Audio-Files) durchsuchen. Konzipiert für den Medienbetrieb ist es nun unter anderem als Datenbanksystem der Apa – Austria Presse Agentur im Einsatz und arbeitet hinter der Recherche-Oberfläche des Apa-Online-Managers (AOM), in dem 8000 User, davon bis zu 1500 gleichzeitig, in den mehr als 80 Mio. Dokumenten aus rund 200 Quellen recherchieren.

Bei der Recherche wird der Benutzer durch mächtige Suchfunktionen unterstützt. Wichtigstes Feature in diesem

Zusammenhang ist das Themen-Clustering. Der Benutzer gibt einen Suchbegriff ein und erhält ein Suchergebnis, das allerdings weit über herkömmliche Suchmaschinen-Listen hinausgeht. Power Search teilt die Trefferliste automatisch in Themen-Cluster auf und gruppiert sie in sogenannte Themencluster. Jeder dieser Cluster wird automatisch mit den wichtigsten Themenschlagwörtern versehen. So findet das System beispielsweise zum Suchbegriff „UNO“ die Themengruppen (Cluster) „Nordkorea/Sanktionen“, „Iran/Ahmadinejad“ oder auch „Israel/Bericht“. Bei Interesse an

einem Themenbereich kann der User auf einen Cluster klicken und sich so durch die Themengruppen navigieren.

Informationsvisualisierung

Wirklich spannend wird das Themen-Clustering in Kombination mit der Power-Search-Informationsvisualisierung. Die Entwickler von Apa-IT haben in diesem Bereich Pionierarbeit geleistet und stellen dem Benutzer zusätzlich zu Trefferlisten verschiedene Möglichkeiten zur bildlichen Darstellung von Suchergebnissen zur Verfügung. Der User kann aus mehreren Visualisierungsvarianten

wählen. Eine Möglichkeit dafür ist die Hügeldarstellung, in der Cluster, die thematisch sehr eng mit dem Suchbegriff verknüpft sind, als größere Hügel, Cluster, die weniger Relevanz für den Suchbegriff besitzen, als kleinere Hügel dargestellt werden.

Neben den Relevanzdarstellungen kann der Benutzer auch den zeitlichen Bezug eines Themas sichtbar machen. Wählt er beispielsweise den Suchbegriff „Internet“ und sucht über mehrere Jahre zurück in den Dokumenten der Apa, erkennt er in der Zeitreihenvisualisierung, dass vor der Mitte der

90er Jahre das Internet als Begriff nicht vorkommt, Anfang der 2000er Jahre einen Boom erlebt und seither nach einem kurzen Abfall relativ stabil in der Berichterstattung und damit in den Dokumenten der Datenbank vorkommt.

Neben Themen-Clustering und Informationsvisualisierung bietet Power Search noch eine Reihe von Funktionen zur Erleichterung der Suche, wie die unscharfe Suche, Wortstamm-suche oder die Profifunktion „Near“, die sich den Abstand zwischen zwei gesuchten Begriffen in Textdokumenten als Kriterium heranzieht. *bra*