



Enterprise Knowledge Retrieval System

Vidoretrieval - Suchmaske

Suche anhand Stichworten/Metadaten

gesuchte Person:

weitere Person

Filmtitel:

Regisseur:

Inhaltsbasierte Suche

bildbasierte Suche aktivieren:

Anfragebild:

Datum:

IRSTREAM

Metadatenverwaltung

Serverbezeichnung:

verwende eingebettete Datenbank

überschreiben

verwende externe Datenbank

Datenbankhost:

Datenbanktyp:

Datenbankname:

Portnummer:

Benutzername:

Passwort:

Die Allround-Suchmaschine für Ihr Unternehmen

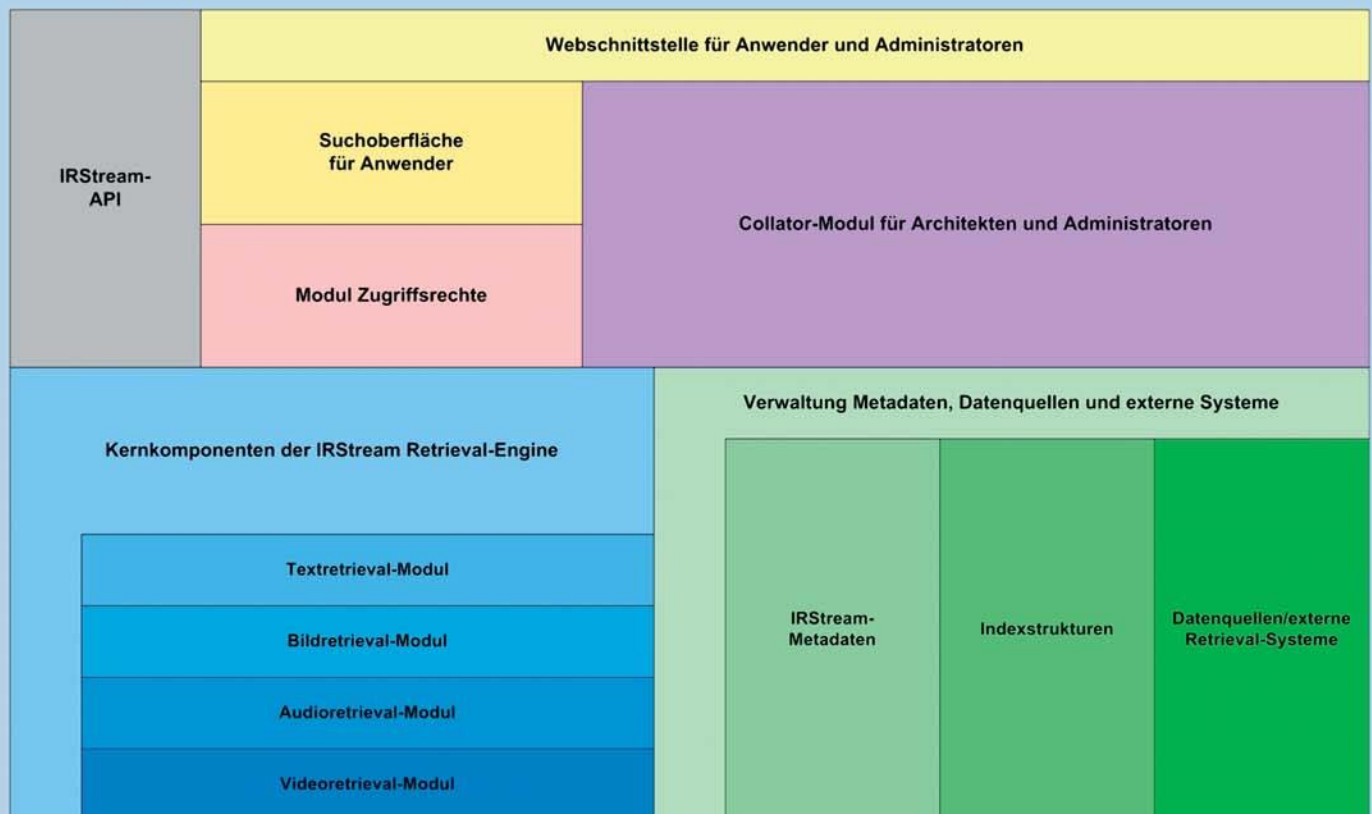
Jeder kennt diese Sprichworte, die in der heutigen Informationsgesellschaft immer mehr zur Realität werden. In den heutigen Datennetzen stehen sehr viele Informationen zur Verfügung, auf die zugegriffen werden kann. Diese Informationen enthalten dabei neben reinem Text, immer mehr auch multimediale Anteile, wie Bilder, Musik oder Videos. Doch aufgrund der Vielzahl sehr unterschiedlicher Arten von Informationen wird es der Informationssuchende bei dieser Entwicklung bald sehr schwer haben, genau die Informationen bzw. das Wissen zu bekommen, die er eigentlich sucht.



IRStream ist eine intelligente Enterprise Search Engine, die zukünftige Anforderungen an die Suche in Unternehmensnetzen abdeckt. IRStream unterstützt hierbei die Suche, nicht nur auf Text sondern auch auf Bildern, Videos und Audio-Dateien. IRStream kann zudem sämtliche vorhandenen Datenquellen nutzen, Internet, Intranet, Datenbanken, Desktop etc. Hierbei kann IRStream bei der Suche sowohl die innere Struktur der Dokumente als auch die Beziehungen (z.B. organisatorisch) zwischen den verschiedenen Informationsquellen berücksichtigen. IRStream verfügt darüber hinaus über ein intelligentes Berechtigungskonzept, so dass Nutzer nur die Daten zu sehen bekommen, für die er geeignete Berechtigungen hat.

Grundlegende Konzepte von IRStream

Im Bereich der Textsuche unterstützt IRStream die Volltextsuche, die phonetische Suche und die inhaltsbasierte Ähnlichkeitssuche, bei der eine Suchanfrage nicht nur nach exakten Treffern sucht, sondern auch ähnliche Wörter/Wortstämme ins Suchergebnis einfließen lässt. Ferner ermöglicht IRStream auch die Suche in der Struktur von Dokumenten. Hierbei kann nicht nur nach bestimmten Dokumentteilen gesucht werden, es ist ebenfalls möglich sich als Ergebnis einer Suchanfrage bestimmte Fragmente (Absatz, Paragraph etc.) eines Dokuments zurückliefern zu lassen. Die inhaltsbasierte Suche auf Bildern, Audiodaten und Videofilmen wird anhand von Merkmalen wie Farbverteilungen, Formgebungen etc. unterstützt. Neu ist zudem, dass IRStream die medienübergreifende Suche innerhalb einer Suchmaschine anbietet. Anders als Mengenorientierte Suchsysteme verarbeitet IRStream Anfragen in Form von Informationsströmen und ermittelt für dem Anwender immer nur die gerade benötigten Ergebnisse (Top-k Anfragen), wodurch die Suche in vielen erheblich effizienter wird. Die Systemarchitektur von IRStream wurde modular gestaltet und ist damit offen für spätere Erweiterungen, siehe folgende Abbildung:



Webschnittstelle für Suchanfragen

Wie schon zuvor beschrieben unterstützt IRStream in einem System die inhaltsbasierte Suche auf unterschiedlichen Medien. Im Folgenden soll exemplarisch die Webschnittstelle für die kombinierte Suche auf Textdokumenten und Bilddaten vorgestellt werden:

The screenshot shows a search interface with two main sections: 'Inhaltsbasierte Textsuche' and 'Inhaltsbasierte Bildsuche'. The text search section includes a text input field with the example text 'Schröder bekommt Job bei Gasprom', a 'Datei...' button, and a list of search options under 'Erweiterte Suchoptionen'. The image search section includes an image input field with a photo of Schröder, a 'Datei...' button, and similar search options. At the bottom, there are dropdown menus for 'Kombinierte Suche' (set to 'im selben Dokument'), 'Auswahl Ergebnisgranulat' (set to 'Text und Bild'), and a 'Suche starten' button.

Im linken Teil der Benutzeroberfläche kann der Nutzer alle Einstellungen für die Textsuche vornehmen. Hier kann er die Informationsquellen (Auswahl Suchbereich) angeben, auf denen gesucht werden soll- Ferner kann er auswählen, welche Form der Suche durchgeführt werden soll (Art der Suche) und wie sein Anfragetext lautet. Des weiteren kann er das Ergebnisgranulat der Suche bestimmen und angeben, in welchem Format die Dokumente vorliegen sollen. Die Konfiguration der Bildsuche läuft analog, wobei für die Suche ein Anfragebild anzugeben ist. Möchte der Nutzer beide Formen der Suche kombinieren, so kann er die Option „kombinierte Text-/Bildsuche“ auswählen und festlegen, in welche Beziehung Text und Bild stehen sollen (keine, im gleichen Dokument, selben Kapitel, usw.) Die dem Nutzer angezeigten Optionen werden dynamisch ermittelt, anhand der für ihn aufgrund seiner Zugriffsrechte bestehenden Möglichkeiten. Die Ergebnisse der Suche werden dem Nutzer aufgeschlüsselt nach Fundort und Medientyp dargestellt. Er kann auf Wunsch Ergebnisse ausblenden sortieren usw., siehe auch folgende Abbildung:

The screenshot shows search results under the heading 'Suchergebnis'. It indicates 'Anzahl Treffer gesamt: 7'. The results are grouped into 'Desktop Anzahl Treffer: 4'. Under 'Textdokumente', there are two results with relevance scores of 97.8 and 91.3, including titles like 'Schröder übernimmt Führungsjob bei russischem Gasriesen'. Under 'Bilddaten', there are two image results with relevance scores of 99.8 and 87.3, including image names like 'schroeder.jpg' and 'images.jpg'.

Videoüberwachung:

Behörden und Unternehmen setzen immer mehr auf Videoüberwachung. Gerade in unsichern Zeiten, wie heute, gewinnt diese Technologie immer mehr an Bedeutung. Allerdings ist die Suche und Auswertung der archivierten Aufnahmen um bestimmte Ereignisse zu suchen immer noch mit hohem Zeitaufwand verbunden. IRStream analysiert schon während der Aufnahme das Filmmaterial. Wenn ein Anwender wissen möchte, zu welchen Zeiten LKW's oder ob ein bestimmter LKW mit der Aufschrift "XY" vorgefahren ist, kann er dies als Sucheingabe eingeben definieren und IRStream liefert dann die entsprechenden Sequenzen als Suchergebnis zurück.

Videoretrieval in Medienunternehmen:

Fernseh- und Rundfunkanstalten benötigen oftmals Filmmaterial sehr kurzfristig, um auf aktuelle Ereignisse zeitnah reagieren zu können. Oftmals wird vorhandenes Filmmaterial für neue Beiträge wieder verwendet. Heute müssen Redakteure mit sehr hohem Zeitaufwand in analogen oder digitalen Filmarchiven nach den relevanten Filmsequenzen suchen. Um einen Beitrag über Bundeskanzler Schröder zu erstellen, kann der Redakteur über die Benutzeroberfläche von IRStream ein Bild des Kanzlers eingeben und zudem weitere Suchbegriffe verwenden, wie z.B. Neujahrsansprache oder Name des Redakteurs, da IRStream auch in der Struktur und den Metadaten des Dokumentes suchen kann. Als Ergebnis erhält der Redakteur Videosequenzen oder komplette Filme nach Relevanz geordnet angezeigt.

The screenshot shows the 'Videoretrieval - Suchmaske' window with the following fields and options:

- Suche anhand Stichworten/Metadaten:**
 - gesuchte Person: Schröder, Gerhard (dropdown)
 - weitere Person (button)
 - Filmtitel: Neujahrsansprache (text input)
 - Regisseur: (text input)
 - Drehort: Berlin (text input)
 - Jahr: 2004 (dropdown)
 - Kategorie: Ansprachen (dropdown)
- Inhaltsbasierte Suche:**
 - bildbasierte Suche aktivieren:
 - Anfragebild: schroeder3.jpg (text input) | Datei auswählen... (button)
 - 
 - Untertitelerkennung verwenden
- Suche auf Registrierungsdaten:**
 - Videoformat: DV-Standard (dropdown)
 - Auflösung: PAL-D (720x576) (dropdown)
 - Länge (min.): 120 Sekunden (text input)
 - Framerate: 25 Bilder/sec. (text input)
- Ergebnistyp festlegen:**
 - komplettes Video zurückliefern
 - nur relevante Videosequenzen zurückliefern
- Buttons: Anfrage speichern..., Anfrage einladen..., Anfrage starten

Bibliotheken:

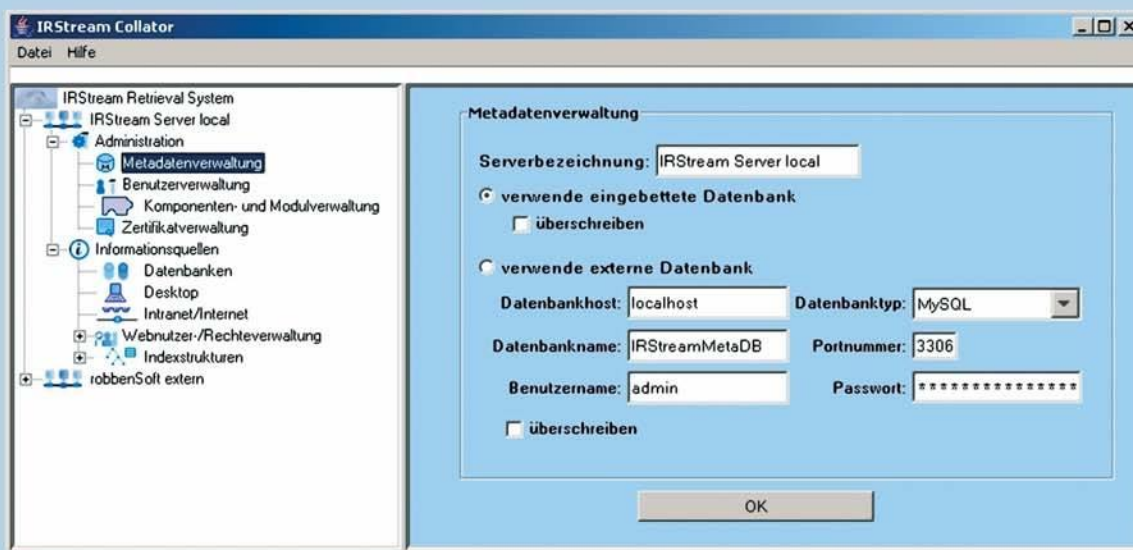
In Bibliotheken und Redaktionshäusern werden Unmengen an Daten elektronisch archiviert und dokumentiert. Eine Suche basiert heute auf einer einfachen Volltextsuche integriert in Dokumentenmanagementsystemen, die eine korrekte Verschlagwortung voraussetzt. Dies ist in der Realität oftmals nicht der Fall. Daher gestaltet sich eine Suche vor allem nach anderen Dokumenttypen oder Teilen, wie z. B. Bildern, sehr schwierig. Da IRStream die Bildsuche oder auch die kombinierte Bild-/Textsuche unterstützt, lassen sich gerade in diesem Marktsegment die Vorzüge von IRStream aufzeigen. Der Anwender kann sowohl ein ähnliches Bild dessen, welches er sucht, als auch Texte, die mit diesem Bild im Zusammenhang stehen könnten, als Suchbegriff eingeben. IRStream liefert dem Anwender dann als Ergebnis eine Rankingliste sortiert nach Relevanz zurück.



Vertragswesen:

In großen Unternehmen, Rechtsanwaltskanzleien oder Behörden werden täglich neue Verträge verfasst. Jeder Vertrag ist einzigartig und doch wiederum so gleich mit 1000 anderen. Mit einer intelligenten Textähnlichkeitssuche kann man leicht eine Rankingliste erzeugen, die dem Anwender eine schier endlose Liste von Ergebnissen anzeigt, wo ähnliche Paragraphen oder Textstellen vorkommen. Diese Suche hat jedoch zwei Probleme: zum einen werden meist nur komplette Dokumente als Suchergebnis geliefert d.h. ein kompletter Vertrag, was juristisch bedenklich sein könnte, und zum zweiten muss dann in dem oftmals mehr als 100-seitigen Dokument, die entsprechend gesuchte Passage selbst erstmal gefunden werden. Bei einer sog. strukturierten Suche tritt dieses Problem nicht auf. Strukturiert heißt, dass zusätzlich zur reinen Textsuche auch anhand der Struktur des Dokumentes gesucht werden kann. Die Struktur ist sozusagen das Inhaltsverzeichnis des Dokumentes und beinhaltet Informationen wie Autor, Erstellungsdatum, Überschriften, Paragraphen und Klauseln. Wenn man zu den gesuchten Begriffen jetzt noch eingibt, dass diese beispielsweise im Paragraph "Gewährleistung" stehen sollen, erhält man mit IRStream lediglich alle relevanten Gewährleistungsparagraphen mit den entsprechenden Begriffen angezeigt. Das heißt, der Anwender bekommt bei der Suche nur relevante Teile von Verträgen zu Gesicht. Zudem kann IRStream bei der Suche die Zugriffsrechte von Personen auf Dokumente mit berücksichtigen, so dass ein Anwender nur Zugriff auf Dokumententeile erhält, die er selbst auch sehen darf.

Der Collator ist das Administrationstool von IRStream. Da unternehmensspezifische Suchanfragen oftmals sehr unterschiedliche Anforderungen muss das Suchprofil vom Administrator konfiguriert werden können, um später dem Anwender optimale Suchergebnisse zu liefern. Außerdem übernimmt der Collator die Benutzer- und Rechteverwaltung, sowie das Management der, für den Betrieb der IRStream Retrieval Engine intern benötigten Metadaten. Ferner kann über den Collator die IRStream Retrieval Engine und der integrierte Applikationsserver verwaltet werden. Hierzu gehören beispielsweise die Integration neuer Module zur Funktionserweiterung oder die Verwaltung von Zertifikaten zum sicheren Datenaustausch zwischen den Web-Clients und IRStream.



Für die optimale Zusammenstellung des Suchraums für unterschiedliche Nutzer, können im Collator verschiedene Datenquellen zu sog. Informationsquellen verbunden werden. Auf diese Weise lassen sich sehr einfach organisatorische Gegebenheiten für spätere Suchanfragen modellieren, z .B. getrennte Informationsquellen für die Abteilungen „Einkauf“ und „Produktion“.

Ein weitere zentrale Aufgaben des Collators ist die Benutzer- und Rechteverwaltung. Die Verwaltung der Benutzer und deren Zugriffsrechte erfolgt auf verschiedenen Ebenen:

- Verwaltung administrativer Logins für Retrieval Engine in interner DB
- Verwaltung der Zugriffs auf Datenquellen (getrennt von Webnutzern) in interner Datenbank
- Verwaltung der Webnutzer über eigenes Modul im Collator, dieses ermöglicht:
 - Verwaltung von Logins und Rollen
 - Konfiguration, welcher Web-Nutzer Zugriff auf welche Informationsquelle bekommt
 - Webnutzer können wahlweise in interner DB, über LDAP, über Active Directory oder externe Datenbanktabellen verwaltet werden

Features

- Einsetzbar unter Windows oder Linux
- Information Retrieval auf Dokumenten unterschiedlichster Art (Text, Bilder, Video, Audio, XML)
- Funktionsumfang Textretrieval:
 - exakte Suche, Ähnlichkeitssuche, phonetische Suche
 - Suche nach Wortgruppen
 - partielle Suche mit Wildcards
 - unterstützte Sprachen: englisch/deutsch
 - Formate: TEXT, XML, HTML, Word, PDF
- Funktionsumfang Bildretrieval:
 - inhaltsbasierte Suche anhand von Farb-, Form- und Texturmerkmalen
 - Suche auf Bildern und Bildregionen (Segmentierung regionen- oder konturbasiert)
 - Formate: JPEG, GIF, PNG, TIFF, BMP
- Funktionsumfang Audioretrieval:
 - Analyse im Zeit-/Frequenzbereich
 - Spektrum, Energie,...
 - Formate: WMA, MP3, AU
- Funktionsumfang Videoretrieval
 - Szenenerkennung
 - Formate: MPEG2, AVI
- Kombinierte Suche: z. B. Text + Bild, dabei auch organisatorische Beziehungen auswertbar
- Mindestanforderungen an Hardware:
 - 1 GB RAM, Prozessor ab 2 GHz, 40 GB freier Platz auf Festplatte
- Suche auf Dokumentenstrukturen z.B. in Kapiteln, Absätzen, Überschriften, Paragraphen
- schnelle Durchführung von sog. Top-k Anfragen mit Hilfe spezieller Indexstrukturen
- externe Suchmaschinen als Datenquellen nutzbar
- verschiedenste Datenquellen nutzbar: Datenbanken, Desktop, Intranet, Internet
- Web-basierte einfache grafische Bedienoberfläche
- Integrierter Applikationsserver, Datenaustausch über XML-Dokumente und SOAP (HTTP und HTTPS)
- Intuitive Administrationsoberfläche zur Vorbereitung und Optimierung von Suchanfragen sowie Verwaltung interner Metadaten
- Umfangreiche Benutzer- und Rechteverwaltung, wahlweise:
 - interne Nutzerdatenbank
 - LDAP/Active Directory-Kopplung
 - oder über externe Datenbank
- vollständig in Java implementiert
- offene Architektur, damit leicht in bestehende Systeme integrierbar
- leicht adaptierbar, dadurch auf Kundenwünsche anpassbar



robbenSoft
Überkumstr. 16
96148 Baunach
Deutschland

Tel.: +49 9544 806120-0
Fax.: +49 9544 806120-9
E-Mail: info@robbensoft.com
<http://www.robbensoft.com>