

multiInfo

multiInfo ist ein „offenes Produkt“.

Es wird ständig erweitert. Jeden Monat wächst es um neue Leistungen.

Derzeit besteht multiInfo aus den Komponenten:

multiEngine: Die **multiEngine** ist in der Lage, über so genannte Agenten, das sind „intelligente“ Java-Programme, jede beliebige Informationsquelle an einem Punkt der Organisation zusammenzuführen. Es entsteht ein Point of Access, an dem die gesamte Information der Organisation zur Verfügung steht.

Über den multiEngine AdministrationsClient wird die Überwachung und zentrale Pflege des Systems sichergestellt. Durch seine Ampellogik zeigt er dem Administrator Schwachstellen auf. Über den AdminClient werden auch die Benutzer gepflegt.

Zudem werden über ihn weitere Agenten, sprich weitere Informationsquellen, durch den Administrator hinzugefügt.

multiInfo: Der **FileCrawler** durchforstet das jeweilige Dateisystem einer Informationsquelle und erstellt eine Statistik über alle gefundenen Dateien und deren Formate.

multiInfo: Der **Extractor** analysiert Textdokumentformate und Dokumentformate mit Metadaten. Dazu gehören u.a. Microsoft Word, Excel, Power Point, ODF-, PDF, RTF und TXT-Formate. Ebenso Emails (msg, eml, eml/mbox) sowie komprimierte Formate (zip, targ, gzip, tgz, bzip2), bewusst – wegen des Datenschutzes – nicht jedoch verschlüsselte Emails oder Dokumente.

Der Extractor extrahiert den textlichen Inhalt einer Datei und führt die Vorverarbeitungsschritte für ein Text Mining durch. Es erfolgt eine syntaktische Verarbeitung mit linguistischer Analyse. Dabei entsteht eine Bibliothek mit so genannten „Machine Learning Algorithmen“. Dieses Verfahren ermöglicht die Extraktion von z.B. Personen-, Orts-, Produkt-, Organisations-, Preis-Bezeichnungen sowie von Segmenten wie Überschriften, Zusammenfassungen, Kommentaren.

multiInfo: Der **Indexer** baut einen Index auf, bei dem jeder Begriff dem zugehörigen Dokument zugeordnet wird. Der Index verweist auf alle Dokumente, die diesen Begriff beinhalten. Der Indexer erstellt eine Index-Datenbank. Dabei ist eine daten-parallele Verarbeitung möglich.

multiCollect: Die Recherche-Anfrage erfolgt über das Recherche-System **multiCollect**. Der Benutzer verwendet seinen ihm bekannten Web-Browser unter Verwendung einer einheitlichen, sehr einfach zu bedienenden Benutzeroberfläche. Diese erfordert kein Training.

Auf dem Rechner des Benutzers wird nichts installiert. Ebenso nicht auf den Servern der Informationsquellen (den Original-Systemen, wie Datenbanken, Data Warehouses, Archiven, Flatfiles usw.). Damit existiert auch keine Update-Problematik.

Der Benutzer erhält aufgrund seiner Recherche eine Kurzzinformation (eine so genannte Trefferliste) sowie die Vollinformation und hat zudem die Möglichkeit, eine eigene *Sammelmappe* (Kollektion) zu erstellen. In der Sammelmappe lassen sich Treffer aus Recherche-Ergebnissen individuell zusammenstellen, sortieren und speichern. Es entstehen *persönliche Themen-Kollektionen*.

Es wird nicht das Originaldokument auf dem Rechner des Benutzers gespeichert, sondern „lediglich“ ein Link auf das Dokument im Original-System sowie ein konfigurierbarer Satz von Metadaten.

Eine Sammelmappe ist damit technologisch gesehen eine Liste von Verweisen und Metadaten, die im XML-Format gespeichert werden. Über die XML-Datei kann der Benutzer jederzeit die Dokumente angezeigt bekommen, d.h. eine einfache Navigation vom Sammelmappen-Eintrag zur Vollinformation ist möglich.

Dadurch, dass lediglich die Verweise und einige Metadaten gespeichert werden, kann der Benutzer sehr große persönlichen Sammelmappen anlegen, ohne dass seine Festplatte nennenswert gefüllt wird. Der Intranet-/Internet-Datentransfer wird minimiert.

In keinem Fall wird das Originaldokument durch den Benutzer verändert.

Der *Datenkranz* – eine definierte Menge von Metadatenfeldern – ist für jede Informationsquelle durch den Administrator konfigurierbar.

Es können optional mehrere Sichten (Views) für jede Sammelmappe einer jeden angeschlossenen Informationsquelle durch den Administrator hinterlegt werden.

Die Benutzeroberfläche ist Ihrem Design angepasst. Optional ist ein deutlich anderes Design möglich. Gerne zeigen wir Ihnen dazu Beispiele.

multiInfo: Die Recherche-Anfrage des Benutzers wird an den **SearchServer** weitergeleitet. Dort wird die *Anfrage in eine „kanonische Form“ umgewandelt*, also z.B. in die Normalform der Wortstämme. Ebenso werden zusammengesetzte Wörter analysiert und zerlegt. Es wird nach Wortstämmen gesucht („finden“, statt „gefunden“). Zusammenhänge können so vom Benutzer besser erkannt werden.

Über den Indexer findet der SearchServer die zur Suche passenden Dokumente.

Die Ausgabe erfolgt über multiCollect im Web-Browser.

Es sind optional sind unter anderem folgende, die Leistung erweiternde Komponenten verfügbar:

multiInfo: Dateiformate-Erweiterung

Neben den standardmäßig unterstützten Dateiformaten sind weitere Dateiformate unterstützbar. Ebenso ist eine weitere Verfeinerung der Unterstützung von Dateiformaten nach Vorgabe der Benutzer möglich, etwa die Unterstützung spezieller, vom üblichen Format-Standard abweichender Formatausprägungen. Die Dateiformate werden entsprechend des Bedarfs integriert.

multiInfo: Grafikdateien-Auswertung

Grafikdateien lassen sich ebenfalls auswerten. Dabei ist auch die OCR-Unterstützung, z.B. für gescannte Papierdokumente, vorhanden.

multiInfo: Audio-, Bild- und Video-Analyse

Der Extractor ist in der Lage, Audioformate, Bildformate und – allerdings sehr aufwändig – auch Videoformate zu durchforsten.

multiInfo: Semantics

Vollautomatische maschinelle Herstellung von Beziehungen und deren Darstellung nach dem Tree-Kernel-Verfahren (X gehört zu ..., Y hält Anteile an ...). So lassen sich u.a. Abhängigkeiten erkennen oder Plagiate, genauso aber auch z.B. Online-Feedbacks bezüglich Produkten bzw. Produkt-Neueinführungen auswerten. Dokumente können mittels der „Support Vector Machine“ entsprechend vorgegebener Merkmale klassifiziert werden (z.B. nach wichtig/unwichtig, gut/schlecht). Die Trefferquote liegt dabei in der Regel zwischen 75 und 95 %, je nach Merkmalen. Es können auch viele Internet-Webseiten parallel durchsucht werden, um festzustellen, ob negative Kommentare bzw. negative Tendenzen feststellbar sind (Image Monitoring sowie Trend Monitoring sowie Gerüchte-Monitoring). Ebenso ist es möglich, Inhalte nach inhaltlichen Erwartungen zu beurteilen (entspricht der Inhalt der Erwartung). Diese Semantics-Funktion wird u.a. von IBM eingesetzt.

multiInfo: HighSpeedIndexer

Mit dieser aufwändigen Logik, die Informationsquellen-spezifisch angepasst wird, lässt sich der Prozess des Indizierens um den Faktor 1000 bis 100.000 beschleunigen – je nach Informationsquellen-Typ. Damit können Indizes dort, wo dies erforderlich ist, fast sekundengenau mitgeführt werden. Jede Veränderung in einer Datei bewirkt quasi eine sofortige Anpassung des Index, wenn der HighSpeedIndexer über die Änderung informiert wird.

multiInfo: zentraler Index

Jede einzelne Informationsquelle lässt sich für sich indizieren. Dabei kann das System aus FileCrawler, Extractor, Indexer/HighSpeedIndexer eingesetzt werden oder/und – je nachdem wie „wertvoll“ die Informationsquelle ist – einfachere Indexierungs-Verfahren wie etwa Lucene oder Google Search Appliance.

Ebenso können semantische Suchmaschinen eingesetzt werden wie Retresco, Fast, Omnifind usw. Um die Ergebnisse der verschiedenen Indexierungs- und Beziehungs-Verfahren zusammenzuführen, lässt sich ein zentraler Index aufbauen. Dieser enthält die Indizes aller indizierten Informationsquellen. Damit existiert ein organisationsweiter Index über die gesamte verfügbare Information.

multiCollect: Tag Clouds und Vorschlagsfunktionen, um die Benutzer durch Such-Hinweise bei ihrer Recherche zu unterstützen.

multiCollect: Speichern von Suchanfragen. Benutzer können ihre „Standardsuchen“ immer wieder aufrufen. Sie können zudem für sich selbst dokumentieren, wie sie gesucht haben.

multiCollect: Sammelmappen für Gruppen (Dossierfunktion)

Dossiers sind zusammengeführte Sammelmappen mehrerer Benutzer zu einem gemeinsamen Thema. Mehrere Benutzer bringen ihre Such-Ergebnisse/Sammelmappen gemeinsam in ein oder mehrere Dossiers zu einem Themenkomplex ein. Dossiers können privat in kleinen Benutzergruppen erstellt werden oder auch für größere Gruppen freigeschaltet werden.

multiCollect: Sammelmappen-Export

Sammelmappen/Dossiers können in Word, Excel oder andere Formate exportiert werden.

multiCollect: Reports

Aus Sammelmappen/Dossiers können Reports generiert werden.

Für jedes angeschlossene System kann hierbei durch den Administrator hinterlegt werden, welche Datenfelder im Report angezeigt werden.

multiCollect: Tiefenrecherche

In einem von einem Benutzer erstellten Dokument können Recherche-Ergebnisse zur weiteren Recherche genutzt werden. Als Ergebnis einer Anfrage werden Suchtreffer (Kurzinformation) angezeigt. Die Tiefenrecherche (Folgerecherche) ermöglicht dem Benutzer, dass er über Links in seinen Suchtreffer-Dokumenten weitere Suchen auslösen kann.

Die Ergebnisse der ursprünglichen Suche bleiben im Suchbaum aktiv. Die Suchhistorie bleibt unverändert.
Anwendungsbeispiel: Spontanes Verfolgen eines Zusammenhang-Hinweises; Dokumente mit der in einem Dokument gezeigten Archivnummer suchen; Dokumente eines anderen Informationssystems mittels einer Identifikationsnummer des gefundenen Dokuments suchen.

multiCollect: Daten-Verknüpfung

Durch Daten-Verknüpfung (so genannte mash-ups) entstehen neue Dokumente, die von ihren Datenquellen her an sich gar nicht in dieser Form zusammenzustellen waren.

multiCollect: Administrator-konfigurierbare Push-Dienste für alle Benutzer

Automatisierte Recherchedienste bewirken, dass voreingestellte Recherchen als Push-Dienste ausgeliefert werden. So können alle Adressaten auf dem gleichen Wissensstand gehalten werden. Die Ergebnisse werden, sobald sie durch Änderung auf einem System entstehen, geliefert (gepusht), ohne dass der Benutzer eine Anfrage starten muss. Der Administrator konfiguriert den jeweiligen Push-Dienst als Veränderungs- bzw. Update-Kontrolle. Allen Benutzern stehen die Ergebnisse dieser regelmäßig automatisch ausgeführten Suchen zur Verfügung.

multiCollect: Administrator-konfigurierbare Push-Dienste für Benutzergruppen

Das sind durch den Administrator konfigurierbare Push-Dienste, die (geschlossene) Benutzergruppen darüber informieren, ob es bezüglich einer definierten Recherche-Anfrage neue Ergebnisse gibt (Veränderungs- bzw. Update-Kontrolle). Den Mitgliedern der Benutzergruppen stehen die Ergebnisse der regelmäßig automatisch ausgeführten Suchen zur Verfügung.

multiCollect: Persönliche Push-Dienste für einzelne Benutzer

Der einzelne Benutzer kann für sich regelmäßig automatisch ausgeführte Suchen hinterlegen. Er erkennt so automatisch Neuerungen am Datenbestand zu für ihn interessanten Fragen.

multiCollect: Push-Dienste mit Benachrichtigungsfunktion über Email oder RSS-Feed

Neuigkeiten werden automatisch an alle Abonnenten des Push-Dienstes versendet. Eine Benachrichtigung ist daher auch ohne das Öffnen von multiCollect möglich – beim RSS-Feed genügt hierfür ein RSS-Reader. Das Öffnen der ausführlichen Information ist direkt aus dem Mail-Client oder RSS-Reader möglich.

multiCollect: Audioplayer

Für Ton-Dokumente ermöglicht der Audioplayer die Wiedergabe. Er erlaubt eine individuell konfigurierbare Anzeige von Metadaten und die Integration der Referenzen (Links zu den Original Audio-Dokumenten) auf die Audio-Dateien in Sammelmappen.

multiCollect: Audioplayer mit EDL-Modul

Diese Funktion ist eine Erweiterung des zuvor beschriebenen Audioplayers.

Sie beinhaltet neben dem Audioplayer die zusätzliche Möglichkeit, Schnittlisten anzulegen. Das bewirkt, dass der Benutzer die für seine Tätigkeit wichtigen Teile einer Audio-Datei zusammenstellen kann. Dabei wird das Original-Material nicht verändert (so genannter Offline-Schnitt). Es entsteht „lediglich“ eine sehr kurze Schnittliste (EDL oder Edit Decision List). Jeder Schnitt kann mit einem eigenen Titel versehen werden. Nur diese EDL wird an ein internes Schnittsystem versendet. Der tatsächliche Schnitt erfolgt im professionellen Schnittsystem (z.B. auch im Workflow mit Final Cut Pro). EDLs können in Sammelmappen gespeichert werden. So können beispielsweise wichtige Abschnitte einer Audio-Datei markiert werden.

multiCollect: Videoplayer

Analog dem Audioplayer ermöglicht der Videoplayer die Wiedergabe von Videos mit individuell konfigurierbarer Anzeige von Metadaten und Integration in Sammelmappen. Möglich ist auch die Anzeige und das Anspringen von Keyframes, wenn diese von der Informationsquelle bereitgestellt werden.

multiCollect: Videoplayer mit EDL-Modul

Hier gilt das Gleiche wie beim Audioplayer mit EDL-Modul, jedoch bezogen auf Video-Dateien. Auch in diesem Fall werden nur die sehr kleinen, wenige Kilobyte großen Schnittlisten-Dateien erstellt und an das Video-Schnittsystem übertragen.

multiCollect: Playlisten-Modul für Audio/Video

Im Playlisten-Modul können gesammelte Audio- bzw. Video-Dateien in der im Playlisten-Modul eingetragenen Reihenfolge abgespielt werden.

multiCollect: Image Viewer (Bildbetrachter) mit Leuchttischfunktion

Dieser Bildbetrachter ermöglicht das kombinierte Betrachten von verschiedenen Bildformaten, aber auch von PDFs, RTFs und Textdateien.

multiCollect: Image Transcoder und Thumbnail-generator mit Leuchttischfunktion

Der Image Transcoder ist ein Bild-Wandler mit Skalier- & Drehfunktionen.

Er generiert Thumbnails und bietet eine Preview-Erzeugung für Bandbreiten-Verringerung bei der Übertragung von Bild-Material. Es werden zunächst die weniger hoch aufgelösten Previews betrachtet und mit diesen gearbeitet, bevor dann das anhand der Previews ausgewählte Original-Material auf den Bildschirm geladen wird. Dies ermöglicht eine enorme Bandbreiten-Einsparung, etwa wenn Material über Internet-Leitungen bezogen oder verteilt werden muss.

multiCollect: Workflows

Mittels Workflows lassen sich beliebige Dokumente in weiterverarbeitende Systeme überführen. Die Workflows stoßen Weiterverarbeitungen an und kontrollieren das Zurückführen der Ergebnisse.

Die Verarbeitungsergebnisse dieser Systeme (z.B. Anwendungen wie SAP) werden wieder in der einheitlichen Benutzeroberfläche des Benutzers dargestellt. Er arbeitet mit seinem Web-Browser und ist damit unabhängig davon, wo er sich auf der Welt befindet.

multiEngine: Benutzerverwaltung mit einfachen Funktionen wie dem Zuordnen von Benutzern zu einer oder mehreren Benutzergruppen.

multiEngine: Erweiterte Benutzerverwaltung

Erweiterung der Benutzerverwaltung zur Kombination von zentralen Benutzerverwaltungen (z.B. AD, LDAP, Datenbanken) mit individuell angepasster interner Benutzerverwaltung.

multiEngine: Ausführliche Administrator-Statistik

Der Administrator kann aktuelle, anonymisierte Nutzungsdaten oder anonymisierte Nutzungsdaten aus der Vergangenheit für einen von ihm definierten beliebigen Zeitraum jederzeit abfragen. Die Statistik kann als XML exportiert und beispielsweise in Excel weiterverarbeitet werden.

Die integrierte Tabellendarstellung erlaubt auch den Export mittels cut&paste-Funktion. Spezielle Klick-Statistiken können konfiguriert werden.

Anwendungsbeispiele: Anzahl der Aufrufe aus bestimmten Datenbeständen, Dauer von Suchanfragen, konkrete (anonyme) Nutzungsdaten.

multiEngine: Wizard für Informationsquellen

Konfigurations-Assistent zur einfachen Zuordnung von im Recherche-Client sichtbaren Quellen zu Benutzern und Benutzergruppen.

multiEngine: Weitere Agenten für den Anschluss von beliebigen Arten von elektronischen Informationsquellen. Es können weit über hundert verschiedene Informationsquellen angeschlossen werden. Die Agenten sind „intelligente“ Java-Programme, die es ermöglichen, relativ schnell Anbindungen an Informationssysteme zu erstellen.

multiEngine: Programmier-Schnittstelle zur Entwicklung eigener Agenten

Das Programming Interface ermöglicht die Anbindung weiterer Informationssysteme mit Hilfe der eigenen Entwickler.

Eine Detail-Beschreibung der meisten Komponenten finden Sie in unserem Download-Bereich unter

<http://info.multibo.com>