

multiSearch

„Wir ertrinken in Information, aber hungern nach Wissen“

John Naisbitt

Mit der multiInfo Base steht dem Benutzer die in der Organisation zugängliche Information jederzeit zur Verfügung.

Er kann die für ihn interessante Information filtern und sammeln. Durch geistige Kombination der Information wird sein Wissen erweitert.

Nun kann die Informationsmenge so groß werden oder einen so komplexen Sachverhalt umfassen, dass sie selbst durch gezielte Informationssammlung nicht mehr analysierbar ist.

Ein Beispiel dafür sind Aktenberge und Rechner-Festplatten, die etwa bei einem Landeskriminalamt in Wirtschaftskriminalitäts-Komplexen eingehen.

Diese riesigen Datenmengen zu durchforsten ist für Menschen kaum möglich. Hier hilft nur „maschinelle Intelligenz“ weiter.

Das gleiche gilt für die vielfältigen und in der Regel sehr umfangreichen Informationsquellen mit Text-, Bild-, Grafik-, Audio- und Video-Dokumenten, die in einer Organisation existieren.

Das multiInfo-System unterstützt den Benutzer technisch mit seinem multiSearch.

multiSearch ist in der Lage, maschinell Beziehungen zwischen Daten, Information und Dokumenten zu erkennen.

multiSearch unterstützt den Benutzer beim Aufbau von Wissen.

Jeder Benutzer benötigt Wissen zur Unterstützung seiner Fertigkeiten. Er braucht sein Wissen und seine Fertigkeiten, um „hochproduktiv“ zu sein.

Der Schlüssel für Wissen ist normalerweise die subjektive Bewertung einer gefundenen Information. Dazu muss der Benutzer

1. wissen, wo er die notwendige Information findet
2. genau die Information finden, die seine Aufgabenstellung unterstützt
3. diese Information sinnvoll verknüpfen.

Die multiInfo Base hilft ihm bei den Punkten 1 und 2: Er hat Zugriff auf die gesamte Information der Organisation (im Rahmen seiner Zugriffsberechtigung) und

er findet aufgrund der Recherche die für ihn exakt zutreffender Information.

Wenn die Informationsmenge jedoch sehr groß wird, ist es für Menschen kaum noch möglich, die Information so zu filtern und zu kombinieren, dass das bestmögliche Ergebnis erzielt wird.

Ein Beispiel ist die Optimierung von Verkaufsstellen. Wenn zehntausende von Verkaufsstellen existieren – etwa Selbstbedienungskästen bei Tageszeitungen – einige davon extrem umsatzstark sind, andere aber überhaupt nicht, obwohl sie in sehr gute Lage liegen ... was kann die Ursache dafür sein?

Es gibt hunderte von Faktoren, die gleichzeitig betrachtet werden müssen. Durch Menschen wäre diese Problem kaum zu lösen.

multiSearch löste es erfolgreich.

Oder denken wir an den Fall der zu untersuchenden Aktenberge-/Festplatten-Menge durch Ermittler der Polizei. multiSearch zeigt den Ermittlern, wo es in diesen riesigen Datenmengen Zusammenhänge zwischen Personen, Orten, Zeiten, Dokumenten, sogar handschriftlichen Anmerkungen auf diesen Dokumenten, mitgeschnittenen Gesprächen usw. gibt. Die Ermittler erhalten genaue Hinweise auf Zusammenhänge. Sie können konkrete Spuren verfolgen.

In jeder Organisation stehen viele Dokumente zueinander in Beziehung.

Alles ist irgendwie miteinander verknüpft.

Erst das Erkennen der Verknüpfung erlaubt die passenden Schlüsse zu ziehen.

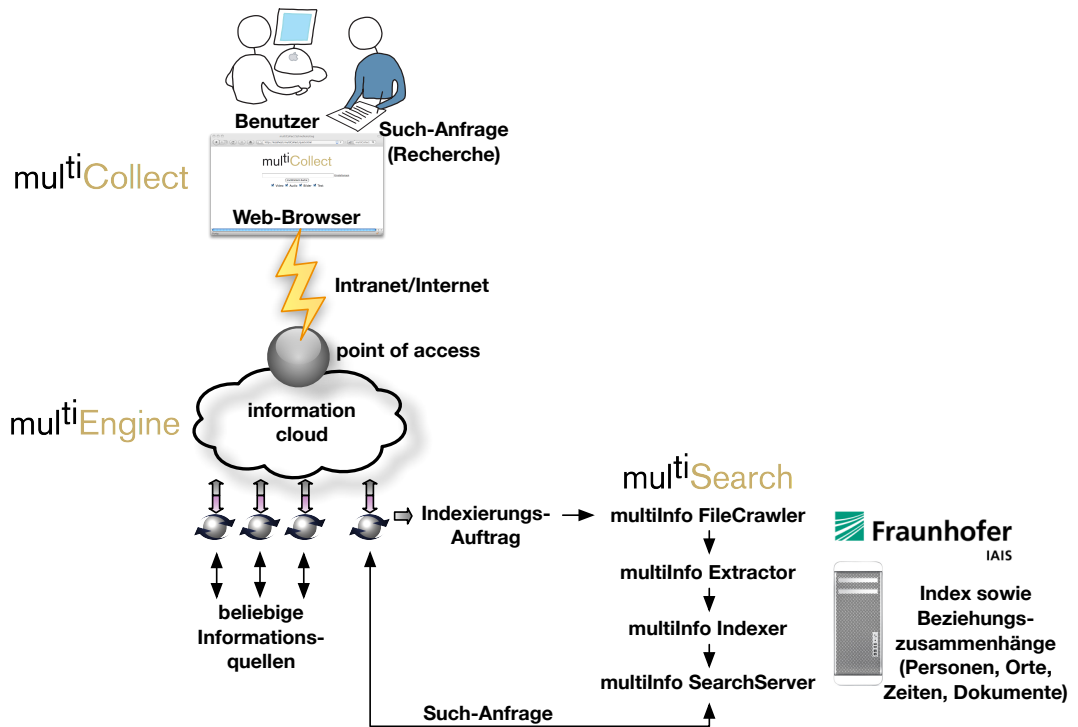
Kein Manager kann Entscheidungen treffen, ohne dass er solche Verknüpfungen anhand von Information vornimmt.

Jeder Experte muss Zusammenhänge erkennen, egal, ob er Techniker ist oder Supportmitarbeiter oder Marketingfachmann, Produktmanager oder Verkäufer oder Finanzexperte.

multiSearch bewirkt, dass jeder nicht nur die Information erhält, die er benötigt, sondern zeigt ihm, wie die Information in Beziehung zu anderer Information steht.

Durch die Kombination herausragender technischer Lösungen und gemeinsam mit den Wissenschaftlern der Fraunhofer-Gesellschaft ist es uns gelungen, ein **einzigartiges Informations-, Indexierungs- und Beziehungszusammenhang-System zu schaffen:**

das multiInfo-System.



Jede beliebige Informationsquelle wird am point of access zusammengeführt.

multiSearch durchforstet die Dateisysteme. Eine Statistik über alle gefundenen Dateien und Formate wird erstellt. Die Dateien werden entsprechend ihren Formaten analysiert. Zu den Formaten gehören u.a. Microsoft Word, Excel, Power Point, ODF-, PDF, RTF und TXT-Formate. Genauso auch Emails (msg, eml, eml/mbox) und ebenso komprimierte Formate (zip, targ, gzip, tgz, bzip2). Aus Geheimhaltungsgründen gezielt verschlüsselte Dokumente bleiben jedoch unangetastet.

Die Vorverarbeitungsschritte für das **Text Mining** werden durchgeführt. Das heißt, die textlichen Inhalte werden extrahiert. Dazu gehören vor allem die „Entities“ (die Funktionseinheiten), sprich Personen, Orte, Dokumente usw. Es ist – durch allerdings aufwändigere Prozesse – auch möglich, Grafiken (etwa Scans), Audio-Inhalte, Bild-Inhalte, ja sogar Video-Inhalte zu analysieren und zu extrahieren.

Aufgrund der Extraktion wird ein Index aufgebaut. Jeder festgestellte Begriff wird dem zugehörigen Dokument zugeordnet. Der Index verweist auf alle Dokumente, die diesen Begriff beinhalten. Es entsteht nun ein neuer physischer Speicher auf einem Rechner, der den kompletten Index der Organisation enthält.

Diese Index-Datenbank ermöglicht statistische Auswertungen. Etwa, wie häufig kommt ein Begriff vor (etwa ein Kundenname) und an welcher Position in einem Dokument.

Diese Index-Datenbank ist die Basis für das auf die Original-Systeme zugreifende **Information Retrieval System**.

Maschinell durchsucht das multiInfo-System alle indexierten Dokumente darauf, ob Beziehungen zwischen den indexierten Begriffen bestehen.

Es lassen sich Schriftverkehr, Emailverkehr, Support, Verträge, Kunden, Lieferanten, Mitarbeiter, Orte, in Beziehung bringen, selbst wenn diese in riesigen Informationsmengen gespeichert sind.

Sucht nun der Benutzer etwas mittels seines Web-Browsers über die einheitliche Benutzeroberfläche der Organisation, wird ihm nicht nur eine Trefferliste bezüglich der Information in den Original-Systemen angezeigt, sondern auch, **wie diese Information mit Information in anderen Informationsquellen im Zusammenhang steht.**

Dazu wird seine Anfrage in eine „kanonische“ Form gewandelt, d.h. in die Normalform der Wortstämme. Es wird nach Wortstämmen gesucht, z.B. „finden“, statt „gefunden“. Ebenso werden zusammengesetzte Wörter analysiert und zerlegt.

Zusammenhänge können damit vom Benutzer besser erkannt werden.

Hinter jedem Stichwort sind die Original-Dokumente einsehbar (wenn die Zugriffserlaubnis dafür besteht).

Kein gespeichertes Dokument geht in der Masse verloren oder wird vergessen.

Die Ausgabe erfolgt über den Web-Browser des Benutzers, egal wo sich dieser befindet.

Es lassen sich nicht nur Dokument-Beziehungen erkennen, sondern auch interessante Kombinationen, etwa welche technischen Zusammenstellungen möglich sind, wer sich bereits mit einem speziellen Thema, für das Hilfe benötigt wird, beschäftigt hat, welche vertraglichen Zusammenhänge bestehen, welche Produkte einem bestehenden Kunden noch angeboten werden sollten usw.

Die gesamte, der Organisation zugängliche Information wird kombinierbar und ist für den jeweiligen Funktionsträger zur Unterstützung seiner Arbeit verfügbar.

Von den hinter seiner Recherche liegenden, hochkomplexen Prozessen bekommt der Benutzer nichts mit.

Er findet das, was er sucht, zutreffender, schneller und komplett und erhält dazu Anregungen für die Lösung seiner Aufgabenstellung.

Einfacher lässt sich Produktivität nicht steigern.