

Metafer – DocumentFinder Rev. 2

Bedienungsanleitung

*Recherchieren
Doubletten finden
Plagiate erkennen
Querlesen
Zusammenfassen
Bewerten*

Bedienungsanleitung

Metafer DocumentFinder
Rev. 2

Metafer GmbH
Mendelstr. 11
48149 Münster

Metafer ist ein registriertes Warenzeichen der Metafer GmbH.

Alle anderen genannten Produkt- und Firmennamen können Warenzeichen oder eingetragenen Warenzeichen der jeweiligen Inhaber sein.

Änderungen der Spezifikation ohne vorherige Ankündigung vorbehalten.

Rev. 2 - Product Code: pd-met-df-de.fm

© Copyright 2000 – 2004 Metafer GmbH Alle Rechte vorbehalten.

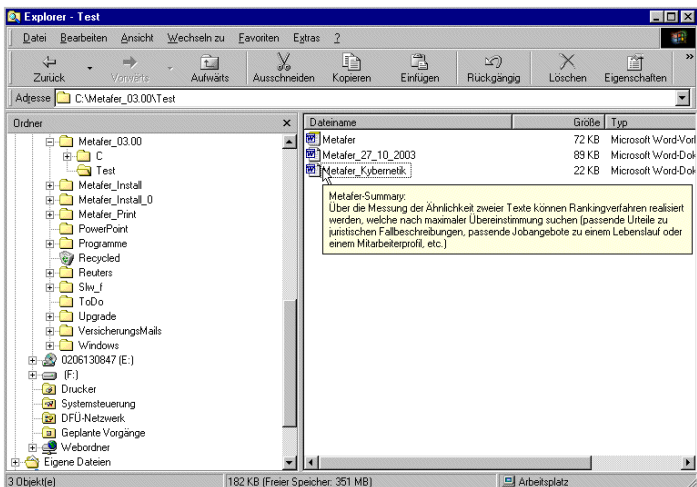
Auflage 1 vom 28.11.2004

Inhaltsverzeichnis

1	Einführung.....	4
2	Installation	5
3	Erfassen des Dokumentenbestandes.....	9
4	Suchen nach ähnlichen Dateien	11
5	Aufruf des Summarizers.....	13
5.1	Dateiformate	15
6	Frequently asked questions.....	17
6.1	Warum wird keine Zusammenfassung angezeigt ?...	17
6.2	Welche Dateiformate werden unterstützt ?	17
6.3	Welche Sprachen werden unterstützt ?	18
7	Known Problems.....	19
7.1	Fehler bei Winword DOC Files.....	19
7.2	Fehler bei PDF Files.....	19
8	Funktionsweise	21

1 Einführung

Der Metafer – DocumentFinder unterstützt Sie bei der täglichen Arbeit durch Zusammenfassungen Ihrer Dokumente und schnelle Recherche nach ähnlichen Dokumenten. Bei Dokumenten mit einem mittleren Volumen (bis zu 300 KByte) wird schon direkt im MS-Windows Explorer, einfach durch Positionieren der Maus auf den Dateieintrag, eine „Ein-Satz-Summary“ angezeigt. Zusätzlich kann über das Kontextmenü zu jeder Datei der „Metafer Summarizer“ direkt gestartet werden. Ebenso kann über das Kontextmenü schnell nach ähnlichen Dokumenten, anderen Versionen des gleichen Dokumentes und themenverwandter Dokumente gesucht werden.

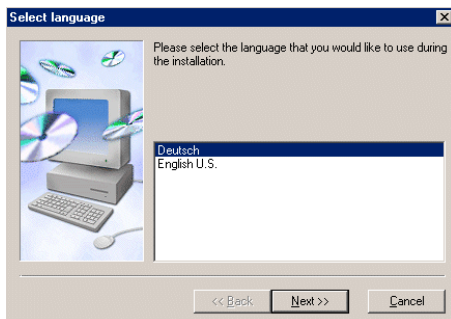


Die Zusammenfassung wird unmittelbar bei Bedarf erstellt. Für die Recherche muss vorher über das Kontextmenü des Windows Explorers der Datenbestand indiziert werden.

2 Installation

Sprachauswahl

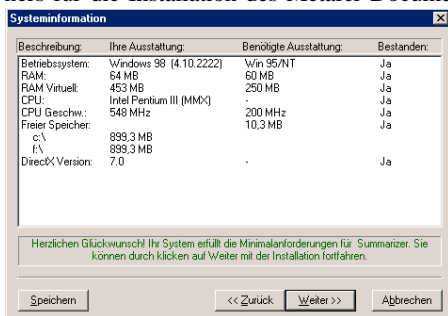
Die Installation des Metafer-DocumentFinders wird gestartet über den Aufruf der „Setup.exe“ von der Installations CD. Nach dem Starten des Installationsprogramms erfolgt zuerst die Aufforderung zur Auswahl der Sprache. Die Installation kann in Deutsch oder Englisch erfolgen.



Markieren Sie die gewünschte Sprache und klicken Sie anschließend auf den „Next-Button“.

Systeminformationen

Das Installationsprogramm prüft nun, ob die Ausstattung Ihres Rechners für die Installation des Metafer-DocumentFinders ausreicht. Das Fenster „Systeminformationen“ zeigt das Ergebnis dieser Prüfung.



In dem Balken unterhalb der Ausstattungsbeschreibung erfahren Sie, ob Ihr System die Minimalanforderungen erfüllt.

Wenn dies nicht der Fall ist, wird mit roter Schrift darauf hingewiesen. Ansonsten sehen Sie in grüner Schrift die Bestätigung und Sie können mit Klicken auf „Weiter“ mit der Installation fortfahren.

Benutzerinformation

Anschließend geben Sie im nun erscheinenden Fenster ihre Benutzerinformationen ein

Benutzerinformationen

Bitte geben Sie unten ihre Benutzerinformationen ein. Benötigte Felder sind mit einem Stern (*) gekennzeichnet.

Name *:

Firma *:

Seriennummer: - -

<< Zurück Weiter >> Abbrechen

- den Namen
- die Firma
- eine gültige Seriennummer.

Falls Sie die zeitlich begrenzte Demo-Version installieren möchten, können Sie diese durch Eingabe der Seriennummer 00000 0000 0000 aktivieren. Die Demo Version hat eine eingeschränkte Laufzeit von 30 Tagen. Nach Eingabe der Seriennummer gelangen Sie über „Weiter“ zur Auswahl des Installationsverzeichnis (Zielverzeichnis).

Zielverzeichnis

Wählen Sie nun das Zielverzeichnis, in dem der DocumentFinder installiert werden soll. Das Zielverzeichnis ist mit einer Standardeinstellung vorbelegt. Möchten Sie den Document-

Zielverzeichnis wählen

Summarizer wird in diesem Verzeichnis installiert.

Klicken Sie Weiter zur Installation in diesem Verzeichnis, Durchsuchen zur Installation in einem anderen und wählen Sie dieses oder geben Sie den Pfad ein. (*Summarizer* wird automatisch hinzugefügt.)

Zielverzeichnis:

Durchsuchen

<< Zurück Weiter >> Abbrechen

Finder dort installieren, klicken Sie einfach auf „Weiter“.

Falls Sie das Installationsverzeichnis abändern möchten, so gelangen



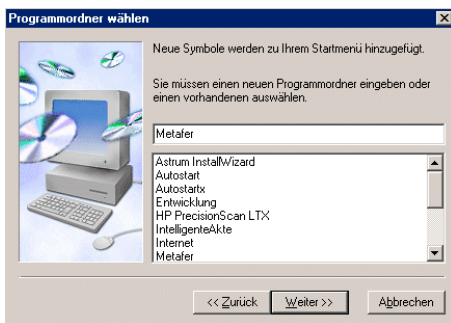
Sie über den „Durchsuchen – Button“ in den Directory Auswahl-dialog und können dort das gewünschte Zielverzeichnis auswählen.

Mit „OK“ bestätigen Sie die Auswahl und gelangen zurück in den Zielverzeichnis Dialog.

Nach Betätigung des „Auswahl-Buttons“ erscheint die Aufforderung zur Auswahl des Programmordners.

Programmordner

Voreingestellt ist der Programmordner „Metafer“ Sie können diese



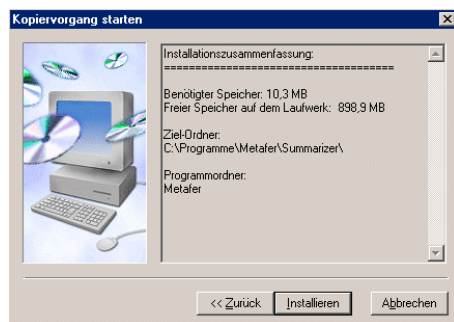
Einstellung übernehmen, oder einen anderen Programmordner auswählen.

Nun haben Sie alle erforderlichen Angaben eingegeben und über den „Weiter-Button“ wird jetzt die Installation gestartet.

Im folgenden Fenster wird Ihnen eine Zusammenfassung Ihrer Eingaben angezeigt.

Kopiervorgang starten

Bereits getätigte Angaben können Sie nochmals abändern, indem Sie



den „Zurück-Button“ anklicken bis Sie zu dem gewünschten Eingabefenster gelangen in dem Sie eine Änderung vornehmen möchten.

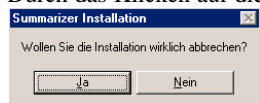
Sind alle Angaben korrekt, starten Sie über „Installieren“ den eigentlichen Installationsvorgang. Ein Fortschrittsbalken zeigt Ihnen den aktuellen Stand des Installationsvorgangs an.

Das Ende der Installation wird über ein entsprechendes Fenster angezeigt.

Installation abbrechen

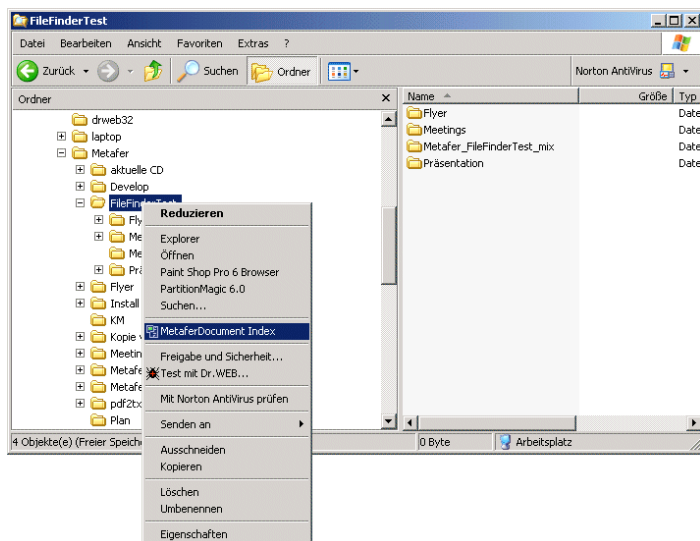
Das Installationsprogramm kann auch jederzeit abgebrochen werden. In jedem Fenster ist ein „Abbrechen-Button“ vorhanden.

Durch das Klicken auf diesen „Abbrechen-Button“, erscheint folgendes Fenster: Wenn Sie den Setup wirklich abbrechen möchten, klicken Sie auf „Ja“. Mit „Nein“ fahren Sie mit der Installation fort.



3 Erfassen des Dokumentenbestandes

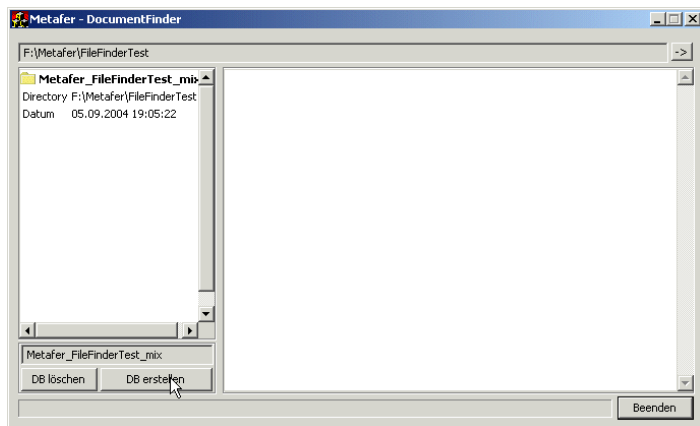
Bevor in einem Bestand von Dokumenten nach ähnlichen Dokumenten gesucht werden kann, muss zuerst mindestens eine Index Datei erstellt werden. Hierzu wählen Sie im Microsoft Windows Explorer ein Verzeichnis aus, welches die Dokumente enthält, die für die Suche berücksichtigt werden sollen. Über das Kontextmenü (rechte Maustaste) kommen Sie zu dem Menüeintrag „Metafer DocumentIndex“.



Über Auswahl des Menüpunktes „Metafer DocumentIndex“ wird der Metafer DocumentFinder gestartet und die entsprechende Datenbank angelegt bzw. falls bereits vorhanden, ausgewählt.

Der Name der Datenbank wird gebildet aus den beiden untersten Verzeichnisebenen. Also zum Beispiel bei der Verzeichnisstruktur:

C:\Metafer\FileFinderTest ergibt sich der Datenbankname: Metafer_FileFinderTest_mdx

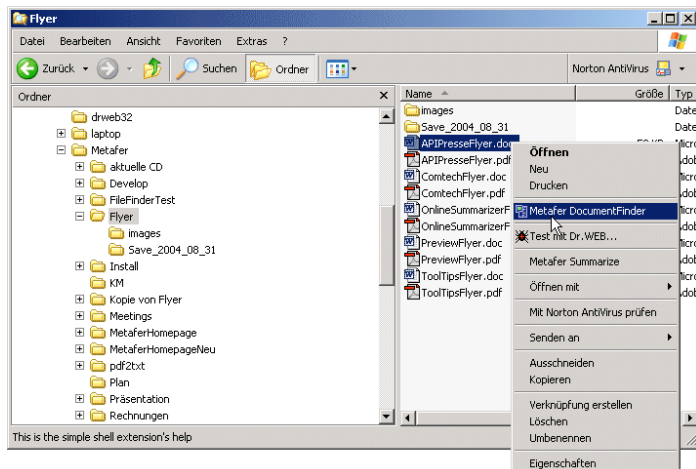


Über den Button „DB erstellen“ wird nun die Datenbank erstellt. Je nach Anzahl der Dokumente und Unterverzeichnisse kann dies von wenigen Sekunden bis zu mehreren Stunden (bei mehr als 50.000 Dokumenten) dauern. Es werden nur Dateien der implementierten Dokumentenformate (HTML, DOC, PDF, TXT, XLS, PPT, C, CPP, H, HPP) berücksichtigt.

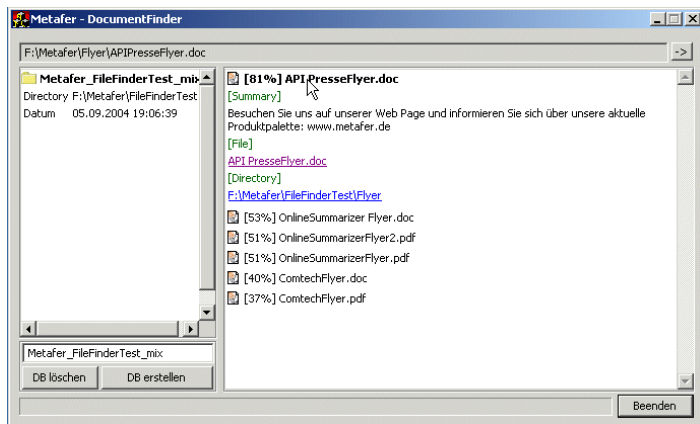
Nachdem die Datenbank erstellt wurde, können Sie das Programm über den Button „Beenden“ beenden und ab sofort die Suchfunktionalität nutzen.

4 Suchen nach ähnlichen Dateien

Im Kontextmenü des Microsoft Windows Explorer zu Dokumenten (HTML, PDF, DOC, TXT, XLS, PPT, C, CPP, H, HPP) können Sie nun über den Eintrag „Metafer DocumentFinder“ im bereits erfassten Bestand nach ähnlichen Dateien suchen.



Das Ergebnis der Suche wird Ihnen unmittelbar angezeigt. Durch Klick auf einen Eintrag in der Ergebnisliste werden Detailinformationen zu dem gefundenen Dokument angezeigt. In der Summary wird ein kurzer Hinweis auf den Inhalt des Dokumentes gegeben und über den Link „File“ können Sie das gefundene Dokument aufrufen. Ausserdem haben Sie die Möglichkeit über den Link „Directory“ sich den Inhalt des Verzeichnisses anzeigen zu lassen in welchem das gefundene Dokument abgelegt ist.

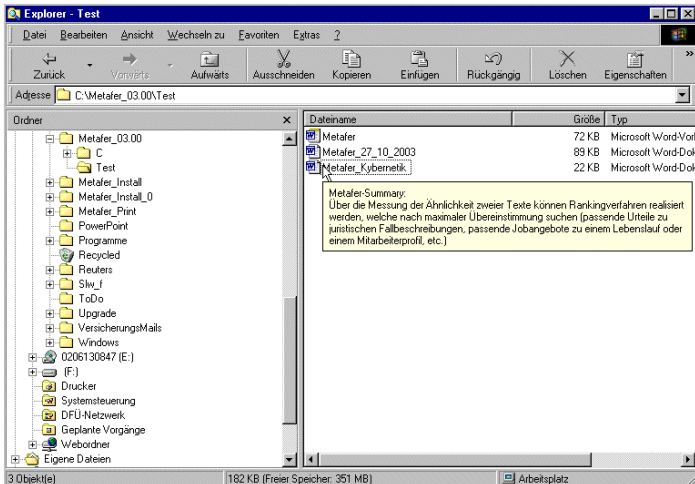


Haben Sie bereits mehrere Datenbanken angelegt, so können Sie im linken Fenster einfach durch Klick auf die entsprechende Datenbank die gleiche Suche in einer der anderen Datenbanken auslösen.

Über die Buttons „DB erstellen“ und „DB löschen“ können Sie jederzeit eine bestehende Datenbank aktualisieren (wenn im entsprechenden Verzeichnis neue Dokumente hinzugekommen sind, oder sich bestehende Dokumente geändert haben) bzw. über den Button „DB löschen“ eine angelegte Datenbank löschen.

5 Aufruf des Summarizers

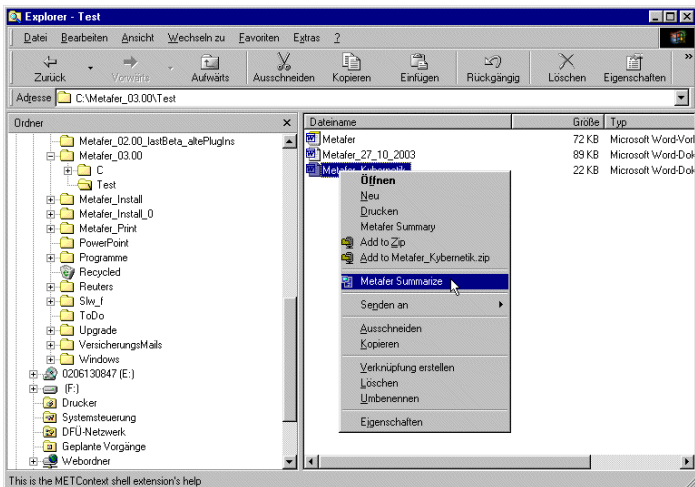
Nach der Installation steht Ihnen die ToolTip-Summarizer Funktionalität unmittelbar zur Verfügung. Für alle unterstützten Dateiformate wird im Microsoft Windows Explorer eine „Zwei-Satz-Summary“ angezeigt, sobald der Mauscursor über eine solche Datei positioniert wird.



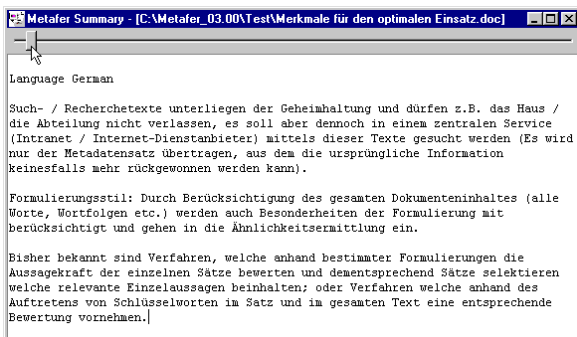
Die Summarizer Funktionalität ist allerdings eingeschränkt auf Dateien bis zu einer Größe von 500 KByte.

Neben der ToolTip-Summarizer Funktionalität steht über das Kontext-Menü des Microsoft Windows Explorers auch eine frei skalierbare Zusammenfassung zur Verfügung.

Durch Auswahl des gewünschten Dokumentes aus der Dateiliste und Anwahl der Funktion „Metafer Summarize“ wird der Metafer Summarizer gestartet und das Ergebnis im Summarizer Fenster angezeigt.



Im Summarizer Fenster wird die aktuelle Zusammenfassung angezeigt und kann natürlich über „Cut & Paste“ in andere Anwendungen übernommen werden.



Der Umfang der Zusammenfassung kann sehr einfach über den Schieberegler oberhalb des Ausgabebereichs verändert werden.



Nachdem die Zusammenfassung erstellt ist, kann durch Verschieben des Reglers nach links oder rechts der Umfang sehr leicht variiert werden.

5.1 Dateiformate

Die unterstützten Formate sind im File Tree entsprechend gekennzeichnet.



Dieses Symbol kennzeichnet unterstützte Dateiformate.



Nicht unterstützte Dateiformate werden über dieses Symbol gekennzeichnet.

Bei der Dokumentenrecherche werden folgende Formate unterstützt:

- DOC Winword DOC Files
- HTM / HTML Files (ohne Frames)
- PDF Acrobat Files
- PPT Powerpoint Files
- XLS Excel Workshets
- TXT reine ASCII Textfiles

Vom Summarizer werden folgende Formate unterstützt:

- DOC Winword DOC Files
- HTM / HTML Files (ohne Frames)
- PDF Acrobat Files
- TXT reine ASCII Textfiles

Nur zu aktuell unterstützten Dateiformaten kann eine Suche durchgeführt bzw. eine Zusammenfassung erstellt werden. Bei HTML Dateien

dürfen keine Frames verwendet sein. MS Word-DOC Dateien und PDF Dateien werden nur verarbeitet, wenn sie nicht Passwort geschützt sind.

6 Frequently asked questions

6.1 *Warum wird keine Zusammenfassung angezeigt ?*

Nur zu den aktuell unterstützten Fileformaten kann eine Zusammenfassung erstellt werden.

Eine Zusammenfassung von PDF-Files kann nur dann erstellt werden, wenn das PDF File weder Password geschützt ist noch besondere Sicherheitseinstellungen (wie beispielsweise die Sperre von "Text/Grafik auswählen") im PDF File aktiviert sind

Der aktuell eingesetzte Winword DOC Konverter liefert unter Umständen fehlerhafte Inhalte wenn in der Winword Datei "alte" eigentlich gelöschte Textbestandteile noch enthalten sind. Überprüfen sie in diesem Falle Ihre Winword Einstellungen und speichern das Dokument erneut über die Funktion "Save-As" von Winword.

Die HTML Datei enthält Frames. Frames werden vom eingesetzten Filekonverter nicht aufgelöst. Daher müssen die eigentlichen Content Seiten für den Summarizer angegeben werden.

6.2 *Welche Dateiformate werden unterstützt ?*

Bei der Dokumentenrecherche werden folgende Formate unterstützt:

- DOC Winword DOC Files
- HTM / HTML Files (ohne Frames)
- PDF Acrobat Files
- PPT Powerpoint Files
- XLS Excel Workshets
- TXT reine ASCII Textfiles

Vom Summarizer werden folgende Formate unterstützt:

- DOC Winword DOC Files
- HTM / HTML Files (ohne Frames)
- PDF Acrobat Files
- TXT reine ASCII Textfiles

6.3 Welche Sprachen werden unterstützt ?

Die Metafer Technologie arbeitet sprachunabhängig. Das System kann jedoch für bestimmte Sprachen und bestimmte Dokumentformate konfektioniert und die Ergebnisse damit optimiert werden.

Die aktuelle Version ist konfektioniert für die Sprachen

- Deutsch
- Englisch

7 Known Problems

7.1 Fehler bei Winword DOC Files

Der aktuell eingesetzte Winword DOC Konverter liefert unter Umständen fehlerhafte Inhalte wenn in der Winword Datei "alte" eigentlich gelöschte Textbestandteile noch enthalten sind. Überprüfen sie in diesem Falle Ihre Winword Einstellungen und speichern das Dokument erneut über die Funktion "Save-As" von Winword.

7.2 Fehler bei PDF Files

Eine Zusammenfassung von PDF-Files kann jedoch nur dann erstellt werden, wenn das PDF File weder Password geschützt ist noch besondere Sicherheitseinstellungen (wie beispielsweise die Sperre von "Text/Grafik auswählen") im PDF File aktiviert sind.

8 Funktionsweise

Der Metafer DocumentFinder wurde auf der Basis der von der Metafer GmbH zum Patent angemeldeten FSR Technologie (fuzzy sense recognition) entwickelt.

Die FSR Technologie beruht auf Verfahren wie sie in der Regelungs- und Automatisierungstechnik zum Einsatz gelangen. Verfahren zur Systemidentifikation im Umfeld hochdynamischer Prozesse sowie Ansätze zur automatischen Strukturierung komplexer Systemabläufe sind sehr gut geeignet auch bei der Ermittlung relevanter Sätze zur Erzeugung von Abstracts, dem Vergleich zweier Dokumente sowie der Suche nach ähnlichen Dokumenten eingesetzt zu werden.

Die Bewertung und Messung von Textähnlichkeiten bzw. dem Grad der Textübereinstimmung lässt sich mit einem schwer beobachtbaren dynamischen System vergleichen. Solange die Software nicht in der Lage ist auf der Basis eines echten Verstehens des Textes zu einer entsprechenden Bewertung zu gelangen – was noch auf Jahre hinaus unmöglich auf jeden Fall aber sehr aufwändig ist – kann eine Bewertung des Textes hoch performant nur auf der Basis einer Vielzahl von indirekten „Meßgrößen“ erfolgen.

Genau hier aber kann mittels entsprechender Methoden (Fuzzy Technologie, Mustererkennung, Erkennung von strukturellen Ähnlichkeiten zur Systemoptimierung, etc.) die Ähnlichkeit und die Relevanz einzelner Systemkomponenten (Texte) ermittelt werden.

Die patentiert FSR-Technologie geht genau von diesem Ansatz aus. Ähnlich wie in schwer beobachtbaren dynamischen Systemen werden eine Vielzahl von indirekten „Messdaten“ erfasst und über das FSR-Verfahren nach Korrelationen gesucht.

Der unschlagbare Vorteil dieses Verfahrens besteht darin, dass neben dem rein sachlichen Inhalt auch Aspekte wie spezielle Dialekte, fachsprachliche Besonderheiten, Formulierungsstil und gefühlsmäßige Färbung des Textes optimal Berücksichtigung finden.

Im Ergebnis führt dies zu einem dem menschlichen Empfinden sehr nahekommenden Ähnlichkeitsmaß.

Mittels einer „Systemidentifikation“ kann vollautomatisch die Besonderheit bestimmter Texte bzw. Kategorien von Texten erfasst und die zum spezifischen Vergleich heranzuziehenden Maßzahlen ermittelt werden. Dergestalt ist es beispielsweise ein Leichtes, die Besonderheit von zum Beispiel juristischen Texten zu identifizieren und ganz allgemein und unabhängig vom konkreten Inhalt nach juristischen Texten zu suchen.

Über die Messung der Ähnlichkeit zweier Texte können Rankingverfahren realisiert werden, welche nach maximaler Übereinstimmung suchen (passende Urteile zu juristischen Fallbeschreibungen, passende Jobangebote zu einem Lebenslauf oder einem Mitarbeiterprofil, etc.)

Ebenso können innerhalb eines Textes die Sätze herausgefiltert werden, welche den Gesamteindruck des Dokumentes am besten wiedergeben (Übereinstimmung bei den signifikanten Messwerten zwischen einzelnen Sätzen und dem gesamten Text). Darüber hinaus besteht sogar die Möglichkeit ein anderes „Referenzsystem“ (Referenztext) zu definieren und dergestalt personalisierte Abstracts zu erzeugen.

Besonderes Kennzeichen dieses Verfahrens ist eine hervorragenden Skalierbarkeit welche einerseits bei sehr hoher Ähnlichkeit selbstverständlich korrekte Ergebnisse liefert und andererseits selbst bei einer nur schwach ausgeprägten Ähnlichkeit (gemeinsamen Systemeigenschaften) immer noch ein korrektes Ranking liefert.

