

Benutzerhandbuch

Teil 1: Anwendung

Inhaltsverzeichnis

1. Was ist InfoCodex?	2
1.1 Möglichkeiten und Mittel von InfoCodex	2
1.2 Typische Anwendungsbeispiele	2
2. Aufrufen von InfoCodex	3
3. Suchen und Finden	4
3.1 Die Informationslandkarte	4
3.2 Retrieval mit Suchtexten	5
3.3 Suchen nach Dokumenteigenschaften (Metadaten)	6
3.4 Anzeige der Ergebnisse	8
3.5 Dokumentfamilien	9
1.1 Alternative Anzeigemöglichkeiten	10
4. Neue Kollektionen anlegen	12
4.1 Kollektionen und Domänen	12
4.2 Eine Kollektion anlegen	12
4.3 Die Eigenschaften einer Kollektion festlegen	12
4.4 Erweiterte Vorgaben für die Kollektionsgestaltung	13
5. Dokumente hinzufügen	15
5.1 Wahl der Datenquellen	15
5.2 Importe ausführen (Inhaltsanalyse/Indexierung)	17
6. Inhalte auswerten	20
6.1 Dokumente kategorisieren	20
6.2 Dokumente hinzufügen	21
6.3 Dokumenteinträge löschen	21
6.4 Teilkollektionen bilden (unter „Inhaltsauswertungen“)	21
6.5 Neuzugänge mit der Basiskollektion vergleichen / Trenderkennung	21
7. Synonyme / Taxonomie	23
7.1 Synonymgruppen	23
7.2 Begriffshierarchie	24
8. System-Management (Kurzfassung von Teil 2 des Handbuchs)	25
8.1 Datenschutz / Domänen-Konzept	25
8.2 Kollektionen verwalten	26
8.3 Beeinflussen der Kategorisierung	26
9. Ausgewählte Beispiele	28
9.1 Informations-Recherchen	28
9.2 Finden von vorhandenem Know-how	28

1. Was ist InfoCodex?

Die "InfoCodex Semantic Engine" ist ein Software-Instrument für das Wissens- und Dokumenten-Management. Es trägt mit seinen Spider-Agents verschiedenste Dokumente aus unterschiedlichen Quellen zusammen (Fileserver, Intra- und Internet, Mailboxen u.a.m.), analysiert die Inhalte der Dokumente über die Sprachgrenzen hinweg, ordnet die Dokumente nach thematischen Gesichtspunkten und stellt den gesamten Inhalt einer Dokumenten-Kollektion in einer übersichtlichen Informations-Landkarte dar – und dies ohne menschliche Intervention.

Erschliessung von Wissen aus unterschiedlichen Quellen

Eine Maschine, die den Menschen beim Durchforsten des Wissensstoffes echt entlasten soll, muss die Dokumente über mehrere Plattformen und Dokumentformate hinweg zusammentragen und entsprechend dem thematischen Inhalt möglichst übersichtlich gruppieren können. Sie sollte auch erkennen, dass die englische Übersetzung eines deutschen Dokuments den gleichen Inhalt hat wie das Original. Die automatische Indexierung für das effiziente Suchen über die Sprachgrenzen hinweg ist unabdingbar.

1.1 Möglichkeiten und Mittel von InfoCodex

<i>Dokumentformate</i>	Spider-Agents für PDF-Files, Word-Dokumente, Excel-Tabellen, PowerPoint-Files, PostScript und EPS, RTF- und Text-Files, HTML, XML, diverse Mail-Formate inkl. Attachments, PST-Files (Outlook-Archive, Persönliche Ordner), JPG, GIF, TIF, BMP, Lotus-Notes, ZIP- und GZIP-Files
<i>Inhaltserkennung</i>	<p>Abstützung auf einer mehrsprachigen linguistischen Datenbank, deren Einträge zwecks Inhaltserkennung mit einer universellen Taxonomie verknüpft sind.</p> <p>Die Datenbank mit derzeit mehr als 4 Millionen Einträgen stützt sich auf bekannte Werke wie WordNet der Princeton University, EuroVoc, AgroVoc, JuriVoc, CIS, DIN und viele weitere Fachvokabulare. Sie verbindet Englisch, Deutsch, Französisch, Italienisch und Spanisch. Russisch, Chinesisch und weitere Sprachen werden durch externe Übersetzungssoftware unterstützt.</p>
<i>Kategorisierung</i>	Semantisches Clustering (mittels linguistischer Datenbank), informationstheoretische Analysen und Neuronale Netze (Kohonen-Map)
<i>Suchindizes</i>	Wörter und zusammengesetzte Begriffe für klassische Volltext-Suche, Synonyme für erweiterte und sprachübergreifende Suche, Ähnlichkeitsmass für Ähnlichkeitssuche (basierend auf dem Neuronalen Netz), z.B. Suche mit freien Textblöcken

1.2 Typische Anwendungsbeispiele

- Suchmaschine für das firmeninterne Netzwerk ("Enterprise Search Engine")
- Informations-Recherchen: Verbindung der Resultate verschiedener Internet-Suchmaschinen mit eigenen Dokumenten, Intranet-Dokumenten, E-Mails u.a.m.
- Marktbeobachtung, Konkurrenzüberwachung, Patentrecherchen
- Wissenskoordination (innerhalb von Projektgruppen und mit externen Stellen)
- Automatische Verschlagwortung und Zusammenfassung von Dokumenten
- Profil-Matching (Vergleich CVs mit Stellenbeschreibungen, Gruppenbildung etc.)

2. Aufrufen von InfoCodex

InfoCodex wird mit einem Standard-Browser bedient (Internet Explorer, Firefox, Chrome).

Startadresse: servername/infocodex.html oder servername/infocodex5.html

In der Startseite erfolgen die Benutzeranmeldung und die Wahl der Dialogsprache (englisch / deutsch / französisch / italienisch / spanisch; für spanisch sind die Masken englischsprachig).

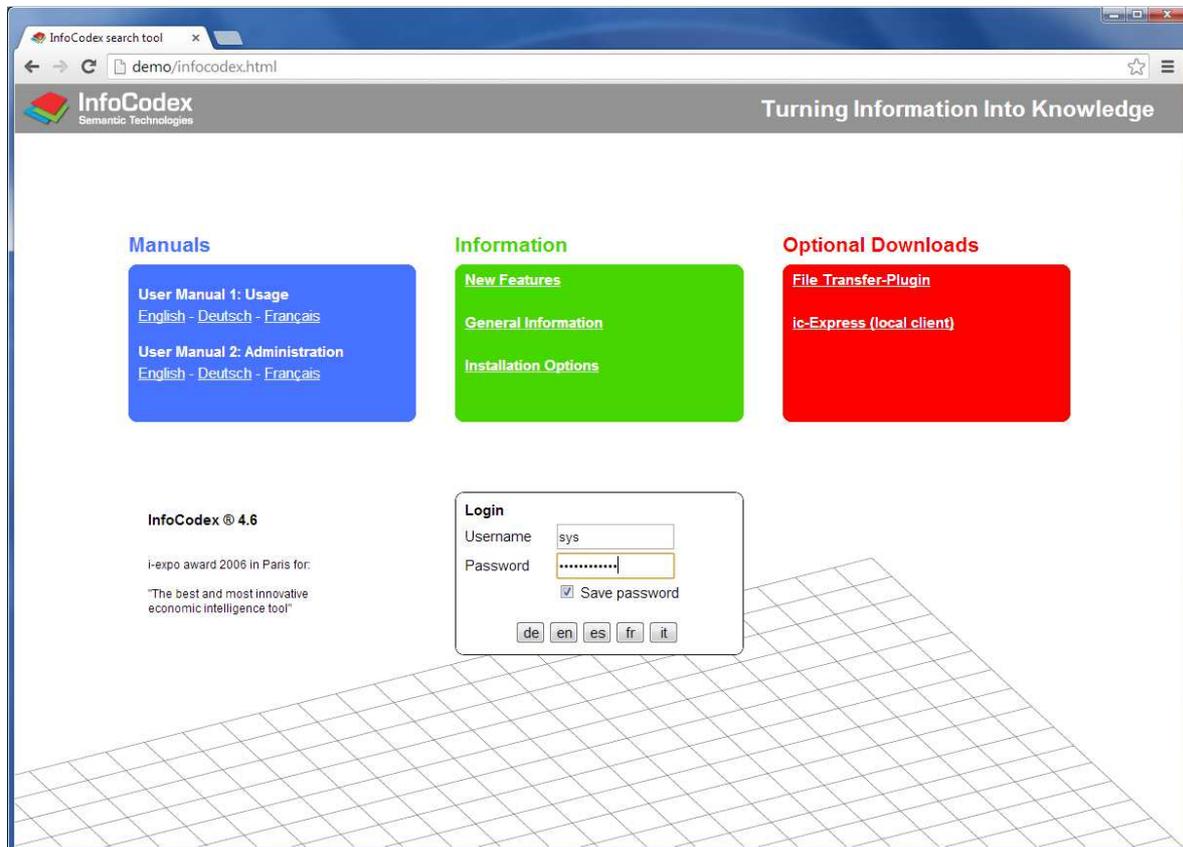


Abb. 1: Startbildschirm

Für das Login gilt je nach Installation:

- Benutzername und Passwort sind gleich wie beim zugrunde liegenden Netzwerk
Dies ist dann der Fall, wenn die Benutzerverwaltung von InfoCodex via LDAP mit einer zentralen Benutzerverwaltung gekoppelt ist ("single point of administration", z.B. Active Directory). Wenn die Zugriffsrechte des Betriebssystems beachtet werden sollen (File System Security), dann ist diese Organisationsform zwingend.
- Benutzername und Passwort sind InfoCodex-spezifisch
In diesem Fall erfolgen die Benutzerverwaltung und die Vergabe von Zugriffsrechten ausschliesslich innerhalb von InfoCodex. Diese Organisationsform erlaubt auch einen User "public" ohne Passwort.
- Automatisierte Anmeldung (keine Passwortabfrage)
Mit der Einrichtung eines "Single Sign-on" entfällt die Abfrage von Benutzername und Passwort.

Nach der Anmeldung erscheint die Standard-Suchmaske, in welcher der Benutzer eine Dokumentensammlung wählen und eine Suche durchführen kann.

3. Suchen und Finden

In InfoCodex wird innerhalb einer Dokumentenkollektion gesucht, welche vorgängig gebildet worden ist. Eine einzelne Kollektion kann Dokumente enthalten, die aus bis zu 10'000 frei wählbaren Quellen zusammengetragen werden, z.B.

- Dokumente aus dem firmeninternen Netzwerk,
- Dokumente aus der Website einer bestimmten Organisation,
- Ergebnisse von Internet-Recherchen mit verschiedenen Suchmaschinen,
- Dokumente von einer CD mit den Referaten einer Fachkonferenz,
- E-Mails und deren Attachments.

Die verschiedenen Kollektionen sind unabhängig voneinander und werden durch InfoCodex einzeln organisiert (d.h. in einem virtuellen Büchergestell dargestellt) und indexiert. Gesucht wird immer innerhalb der ausgewählten Kollektion.

Das Anlegen neuer Kollektionen wird in Kapitel 4 erklärt. Zunächst nehmen wir an, dass bereits einige Kollektionen vorliegen, und es wird vorerst beschrieben, wie die Dokumente gesucht und gefunden werden.

Zuerst wird eine Kollektion aus der Liste links oben im Fenster ausgewählt. Dabei wechselt die Informationslandkarte auf der rechten Seite, um den thematischen Inhalt der Kollektion wiederzugeben.

3.1 Die Informationslandkarte

Die Informationslandkarte gibt eine grafische Übersicht über den Inhalt der gesamten Dokumentenkollektion, vergleichbar mit einem Büchergestell, in dem die Dokumente nach inhaltlichen Gesichtspunkten logisch eingeordnet sind.

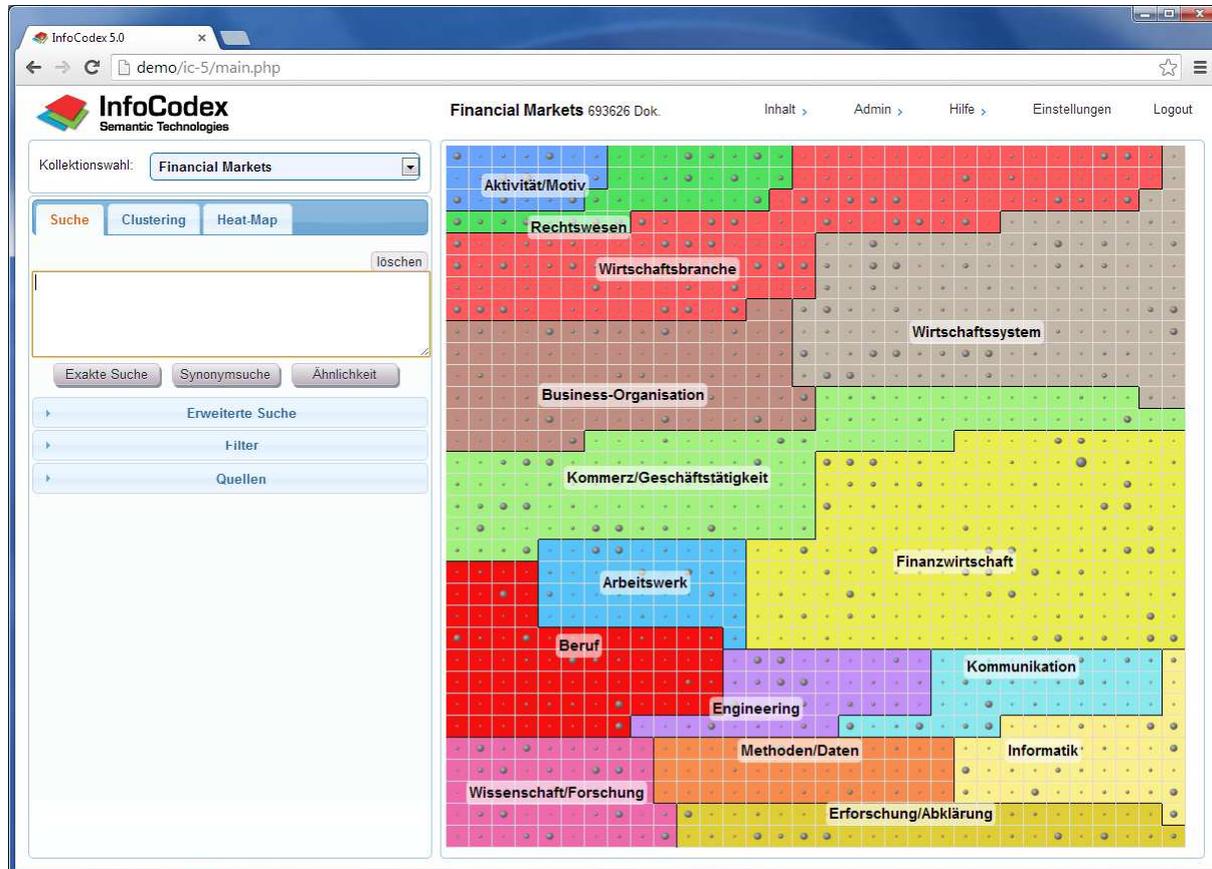


Abb. 2: Suchmaske (links) mit Informationslandkarte (rechts)

In der Landkarte entspricht jedes farbige Feld einem Hauptthema. Jedes Thema besteht aus einem oder mehreren Feldern („Neuronen“), in denen sich wiederum inhaltlich ähnliche Dokumente befinden. Die Grösse des Punktes in der Mitte jedes Neurons gibt Auskunft über die Anzahl Dokumente in diesem Neuron.

Bleibt man mit der Maus über einem Neuron stehen, werden weitere Informationen angezeigt, z.B. Stichwörter zu Themenschwerpunkten und die Anzahl Dokumente. Durch Klick auf ein Neuron kann man die darin enthaltenen Dokumente anzeigen lassen.

3.2 Retrieval mit Suchtexten

Um nach Text zu suchen, gibt man die Suchbegriffe in deutscher, englischer, französischer, italienischer oder spanischer Sprache in das Suchfeld ein. Anschliessend startet man die Suche mit einem der drei Buttons:



Die Suchbegriffe müssen genau in der eingegebenen Form auftreten

Sprachübergreifende Suche nach Synonymen: eine Suche nach „Velo“ findet auch „Fahrrad“, „Drahtesel“, „bicycle“, „bicicletta“ usw.

Es werden inhaltlich ähnliche Dokumente gefunden; eine Übereinstimmung mit den Suchbegriffen ist nicht nötig

Exakte Suche

Die exakte Suche entspricht der klassischen Volltextsuche, wie man sie von den gängigen Suchmaschinen wie Google, Yahoo, Bing etc. kennt. Es wird eine genaue Übereinstimmung der Suchbegriffe mit den Wörtern in den Dokumenten gefordert, wobei zwischen Gross- und Kleinschreibung nicht unterschieden wird. Mit "Velo" wird weder "Velos" noch "Fahrrad" noch "bicycle" usw. gefunden. Bei der Suche werden auch Zahlen und nicht-signifikante Wörter wie "der", "das" etc. berücksichtigt.

Zusätzlich ist auch der Stern als Wildcard am Wortende zugelassen: eine Suche nach "Finanz*" findet alles was mit "Finanz" beginnt, also z.B. "Finanz", "Finanzplan", "Finanzanalyse" usw.

Falls nur wenige Dokumente gefunden werden, bietet InfoCodex an, auch Dokumente zu zeigen, in denen einer der Suchbegriffe fehlt. Die entsprechenden Dokumente werden dann schwächer gewichtet als die diejenigen, die alle Suchbegriffe enthalten.

Synonymsuche

Die Synonymsuche ist eine sprachübergreifende, semantische Suche. Im Gegensatz zur exakten Suche kann sie die Suchbegriffe intelligent interpretieren:

- Synonyme werden erkannt:
Eine Suche nach „Fahrrad“ findet auch die Begriffe „Velo“, „Fahrräder“, „Drahtesel“ usw.
- Es wird sprachübergreifend gesucht:
Mit „Fahrrad“ werden auch „bicycle“, „bicyclette“, „bicicletta“, „velocipede“, „push-bike“ usw. angesprochen.
- Zusammengesetzte Ausdrücke werden erkannt:
„Europäische Union“, „common sense“, „pomme de terre“

Ähnlichkeitssuche

Bei der Ähnlichkeitssuche brauchen keine Begriffe aus dem Suchtext mit den Begriffen übereinzustimmen, die in den gesuchten Dokumenten vorkommen. Die vorhandenen

Dokumente werden mit der Suchanfrage inhaltlich verglichen und nach Ähnlichkeit rangiert. Der Suchtext besteht typischerweise aus einem längeren Prosatext, der z.B. aus einem anderen Dokument kopiert wurde. Der thematische Vergleich und das Ranking erfolgen mithilfe des Ähnlichkeitsmasses von InfoCodex.

Typische Anwendungen der Ähnlichkeitssuche:

- Patent-Recherchen
- Planung von Forschungsarbeiten: Vergleich neuer Forschungsvorhaben mit Arbeiten in der bestehenden Dokumentensammlung
- Vergleich von fremden Publikationen mit den eigenen Arbeiten
- Bearbeitung von Kundenanfragen: Finden ähnlicher Fälle aus der Vergangenheit
- Finden von Normen und gesetzlichen Vorschriften zu einer gegebenen Problemstellung

Eine reine Ähnlichkeitssuche mit wenigen Stichwörtern ohne zusätzliche Einschränkungen ist in der Regel wenig sinnvoll, weil der kurze Suchtext für die Bestimmung der inhaltlichen Ähnlichkeit zu wenig aussagekräftig ist. In solchen Fällen schlägt InfoCodex stattdessen eine Synonymsuche vor.

Und/oder-Verknüpfungen

Von den Wörtern, die durch ein Leerzeichen oder ein Komma getrennt sind, müssen alle vorkommen (UND-Verknüpfung wie z.B. bei Google). Mit ";" (Strichpunkt) können ODER-Gruppen eingeleitet werden.

Suchanfrage	Bedeutung
Verkehr Auto	Sowohl "Verkehr" als auch "Auto" müssen vorkommen
Verkehr; Auto	Entweder "Verkehr" oder "Auto" muss vorkommen
Verkehr Auto; Transport	Entweder die beiden Begriffe "Verkehr" und "Auto" müssen vorkommen, oder der Begriff "Transport" für sich allein.

Bei der Ähnlichkeitssuche machen diese Verknüpfungen keinen Sinn und werden nicht berücksichtigt.

Boolesche Suche

Mit der Booleschen Suche können beliebig komplexe Suchanfragen spezifiziert und mit einer Ähnlichkeitssuche kombiniert werden. Die Regeln und Beispiele sind über den Menüpunkt

Hilfe > [? Erläuterungen zur Suche](#) abrufbar.

3.3 Suchen nach Dokumenteigenschaften (Metadaten)

Neben dem Dokumentinhalt können Dokumente auch nach ihren Eigenschaften gefiltert werden. Man nennt diese Informationen Metadaten.

Erweiterte Suche

Mit der erweiterten Suche kann nach Metadaten der Dokumente gesucht werden. Darunter fallen Eigenschaften wie Autor, Titel, Herkunft und Datum des Dokuments. Ausserdem sind der Dateityp, die ermittelte Sprache und das Importdatum als Metadaten abgelegt.

Eine Einschränkung ist auch via Dateinamen möglich, wobei die Wildcard-Zeichen "*" am Anfang und am Schluss nicht nötig sind, d.h. "proj*vermess" ist gleichbedeutend wie "**proj*vermess**".

Falls beim Import der Dokumente weitere Merkmale wie Kundennummer oder Branche mitgegeben wurden, kann auch nach diesen Merkmalen gesucht werden. Solche zusätzlichen Metadaten werden defaultmässig im Feld "Kommentar" abgelegt (ausser wenn massgeschneiderte Zusatzfelder eingeführt werden).

Filter

Mit Filtern kann das Suchergebnis weiter eingeschränkt werden. InfoCodex erstellt automatisch Filter zu prominenten Themen, Personen, Organisationen und Orten in den analysierten Dokumenten. So kann man die Suche zum Beispiel auf Dokumente beschränken, die von John F. Kennedy und Berlin handeln.

Die Orts-Filter können optional hierarchisch angeordnet werden; im nebenstehenden Beispiel ist „Berlin“ ein untergeordnetes Element von „Deutschland“. In diesem Fall findet eine Einschränkung auf „Deutschland“ automatisch auch „Berlin“. Für die Bildung dieser Hierarchie muss jedoch ein spezielles City/Länderverzeichnis aktiviert werden.

Je nach Kollektion können auch weitere, benutzerdefinierte Filter zur Verfügung stehen. Die

Zahl (57) hinter einem Namen heisst "die entsprechende Ort hat die 57-höchste Relevanz in dieser Kollektion (von den 100 Orten mit den höchsten Relevanzen)".

Quellen

Mit dieser Auswahl lässt sich das Suchergebnis auf bestimmte Quellen – die Herkunft der Dokumente – einschränken.

Eine Kollektion enthält typischerweise Daten aus verschiedenen Quellen wie lokale Dateien aus verschiedenen Netzwerkspeicherorten, Webseiten, Anfragen bei Suchmaschinen usw. Es werden alle Dokumente aus den angegebenen Datenquellen in die Kollektion aufgenommen, soweit InfoCodex die gefundenen Dokumente analysieren kann. Die Herkunft der Dokumente kann beliebig sein; sie hat keinen Einfluss auf die durch InfoCodex gebildete Kategorisierung.

3.4 Anzeige der Ergebnisse

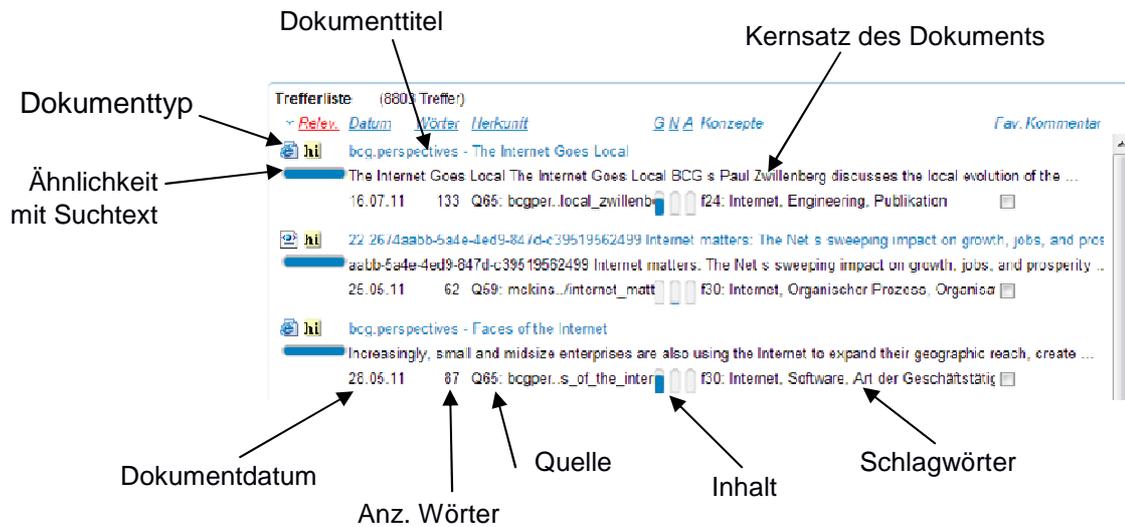


Abb. 3: Detailansicht der Trefferliste

In der Trefferliste können folgende Informationen zu einem Dokument aufgerufen werden:

Detaillierte Dokumentinformationen

Durch Klick auf den Dokumententitel werden weitere Informationen angezeigt:

- Der *Abstract* ist eine automatisch generierte Zusammenfassung der Sätze im Dokument, welche die Kernaussagen des Dokuments am besten wiedergeben.
- Die *Deskriptoren* sind bis zu 18 Stichwörter, mit denen das Dokument thematisch eingeordnet werden kann (angezeigt in der Dialogsprache)
- Im *Kommentarfeld* kann ein beliebiger Freitext zum Dokument hinterlegt werden.



Abb. 4: Dokumentinformationen mit Abstract und Deskriptoren

Originaldokument

Durch Klick auf das Dateisymbol links vom Titel wird nach Möglichkeit das Originaldokument angezeigt. Ist das nicht möglich, z.B. weil es sich um flüchtige Informationen handelt oder das Dokument nicht als einzelne Datei vorliegt (z.B. E-Mail), wird automatisch ein Ersatzdokument mit demselben Inhalt erzeugt und zum Download angeboten.

Textversion mit Highlighting und Visualisierung der Ähnlichkeit

Durch Klick auf das Symbol **hi** wird eine reine Textversion des Dokuments aufgerufen, in dem die gefundenen Suchbegriffe farblich hervorgehoben sind. Durch Klicken des blauen Ähnlichkeitsbalkens unterhalb von **hi** wird die Ähnlichkeit des Dokuments mit der Suchabfrage durch farbige Hinterlegung bzw. farbige Schriften angezeigt. Diese Darstellung bietet einen schnellen Überblick über die relevanten Abschnitte eines Dokuments. Ausserdem wird nachvollziehbar, wie InfoCodex dieses Dokument im Hinblick auf die einzelnen Suchbegriffe bewertet hat.

The screenshot shows the InfoCodex search results for 'Pharma'. A document titled 'Medicine-Worldwide: Malaria' is selected. A 'Marking' window is open, showing a table of related terms and their similarity scores:

Ihre Abfrage:	Dokument:
Wichtigste Begriffe	>> Oberbegriffe
Malaria	>> infection
Impfschutz	>> immunization
	>> medical care
	>> Konzepte

The document preview shows the text 'Medicine-Worldwide: Malaria' and 'Medicine-Worldwide: Malaria' with a similarity score of 0.95. The text is highlighted in yellow, and the search terms are highlighted in red. The document content includes information about malaria, its symptoms, and treatment options.

Abb. 5: Visualisierung der Ähnlichkeit mit der Suchabfragen

3.5 Dokumentfamilien

InfoCodex kann Familien von inhaltlich sehr ähnlichen Dokumenten identifizieren, z.B.

- leicht geänderte oder ergänzte Versionen eines Dokuments,
- dasselbe Dokument in verschiedenen Formaten (Word, PDF, ...), oder
- dasselbe Dokument in verschiedenen Sprachen.

In der Trefferliste erscheint nur ein Dokument pro Familie. Mit einem speziellen Symbol wird markiert, dass mehrere sehr ähnliche Dokumente vorliegen.

The screenshot shows the InfoCodex search results for 'Accelergy Corporation'. A document family is identified, and a 'Document family' window is open, showing a list of documents with their similarity scores and file names:

Document family	Documentenfamilie (fast gleiche Dokumente)	
12.07.11	805	de:ish RegionalAggregateData39Q109Q3Final.xls PricewaterhouseCoopers/National Venture Capital Association I
12.07.11	811	de:ish RegionalAggregateData39Q109Q4Final.xls PricewaterhouseCoopers/National Venture Capital Association I
12.07.11	809	de:ish RegionalAggregateDataQ12010Final (2).xls PricewaterhouseCoopers/National Venture Capital Association I
12.07.11	814	de:ish RegionalAggregateDataQ2 2010_Final.xls PricewaterhouseCoopers/National Venture Capital Association M
12.07.11	807	de:ish RegionalAggregateDataQ32010.xls PricewaterhouseCoopers/National Venture Capital Association MoneyT

Dokumentfamilien haben einen doppelten Nutzen:

- Die Trefferlisten werden kürzer und übersichtlicher.
- Bei Aufräumaktionen können Dubletten bestehender Dokumente einfach identifiziert werden.

1.1 Alternative Anzeigemöglichkeiten

Liefert eine Suchanfrage sehr viele Ergebnisse, kann eine einfache Trefferliste unübersichtlich werden. In diesem Fall stehen zwei weitere einfache Möglichkeiten zur Verfügung, das Suchergebnis thematisch einzugrenzen:

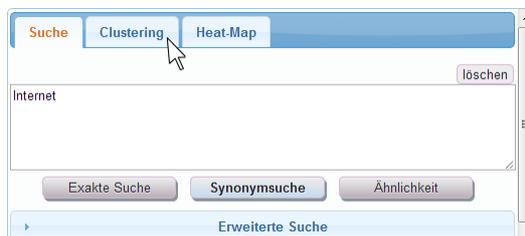


Abb. 6: Weitere Anzeigemöglichkeiten

Clustering

Im Clustering werden alle gefundenen Dokumente nach Themengebiet sortiert in einer Begriffshierarchie aufgelistet. So lässt sich die Trefferliste schnell und einfach weiter thematisch einschränken.

 A screenshot of the InfoCodex 5.0 web application. The browser address bar shows 'demo/ic-5/main.php#'. The page title is 'Financial Markets 693626 Dok.'. On the left side, there is a sidebar with a 'Suche' tab selected. Below the tabs is a table showing a clustering hierarchy. The main content area displays a list of search results for 'Financial Markets' (8803 Treffer). The results are sorted by relevance and include columns for 'Relev.', 'Datum', 'Wörter', 'Herkunft', 'Konzepte', 'Fav.', and 'Kommentar'. The first result is 'bcg perspectives - The Internet Goes Local' with a date of 16.07.11.

N Dok.	Thema	Schwerpunkte
8803	(gesamte Selektion; nur 5000 angezeigt)	
351	Aktivität/Motiv	Zusammenfassung, Erkennungsprozess, /
435	Rechtswesen	Schuld, Finanz, Ort/Gebiet, Preis
1337	Rechtswesen	Finanz, Schuld, Kommunikation
299	Wirtschaftsbranche	Unglück/Verbrechen, Organisationsname,
87	Wirtschaftssystem	Konsument, Internet, Zentrale Dienste, Ort
415	Wirtschaftssystem	Ökonomische Struktur, Finanzwirtschaft, I
418	Kommerz/Geschäftstätig	Organisationsname, Markt, Erwerb
248	Informatik	Internet, Software, Unglück/Verbrechen, O
21		Internet, Engineering, Einkommen, Organi
22		Internet, Engineering, Organisationsname,
39		Internet, Konsument, Ökonomische Strukt
94		Internet, Unglück/Verbrechen, Methode, O
72		Software, Organisationsname, Markt, Recl
493	Informatik	Software, Unglück/Verbrechen, Internet, O
153	Others	Unglück/Verbrechen, Auskunft, Internet, E
764	Others	Arbeitswerk, Engineering, Unglück/Verbre

Abb. 7: Im Clustering (links) sind die gefundenen Dokumente nach Themen aufgegliedert

Heatmap

Die Heatmap entspricht der Informationslandkarte (vgl. Kapitel 3.1), wobei die einzelnen Felder entsprechend ihrer Relevanz eingefärbt werden. Mit der Heatmap kann man sich auch über sehr grosse Mengen gefundener Dokumente schnell einen Überblick verschaffen. Zusätzlich zeigt sie auf, welche Themengebiete innerhalb der Kollektion wie viele relevante Dokumente enthalten.

The screenshot displays the InfoCodex 5.0 web application. The browser address bar shows 'demo/ic-5/main.php#'. The page title is 'Financial Markets 693626 Dok.'. The interface includes a navigation menu with 'Inhalt', 'Admin', 'Hilfe', 'Einstellungen', and 'Logout'. On the left, there is a 'Kollektionswahl:' dropdown set to 'Financial Markets' and buttons for 'Suche', 'Clustering', and 'Heat.Map'. The heatmap visualizes thematic clusters with labels such as 'Wirtschaftsbranche', 'Wirtschaftssystem', 'Business-Organisation', 'Kommerz/Geschäftstätigkeit', 'Arbeitswerk', 'Finanzwirtschaft', 'Beruf', 'Engineering', 'Kommunikation', 'Methoden/Daten', 'Informatik', 'Wissenschaft/Forschung', 'Erforschung/Abklärung', 'Rechtswesen', and 'Aktivität/Motiv'. On the right, a 'Trefferliste (8803 Treffer)' is shown with columns for 'Relev.', 'Datum', 'Wörter', 'Herkunft', 'Konzepte', 'Fav.', and 'Kommentar'. The list contains several entries with blue 'hi' icons, including titles like 'The Internet Goes Local', 'The Net's sweeping impact on growth, jobs, and prosperity', and 'The Future of the Internet'.

Abb. 8: Die Heatmap (links) bietet einen schnellen Überblick über die thematischen Schwerpunkte der gefundenen Dokumente

4. Neue Kollektionen anlegen

4.1 Kollektionen und Domänen

InfoCodex organisiert die indizierte Dokumente in sogenannten **Kollektionen** (vgl. Kapitel 3 und 8.1). Die Dokumente selbst werden beim Hinzufügen zu einer Kollektion nicht verändert oder verschoben, sondern lediglich gelesen und kategorisiert. Für die Verknüpfung mit dem Originaldokument zwecks Sichten wird in InfoCodex nur der Link (URL) gespeichert.

Die Kollektionen können in InfoCodex verschiedenen **Domänen** angelegt werden. Eine Domäne ist eine Sammlung von Kollektionen und dient vor allem der Trennung verschiedener Benutzergruppen und der Vergabe von Benutzerrechten bei hochsensiblen Daten.

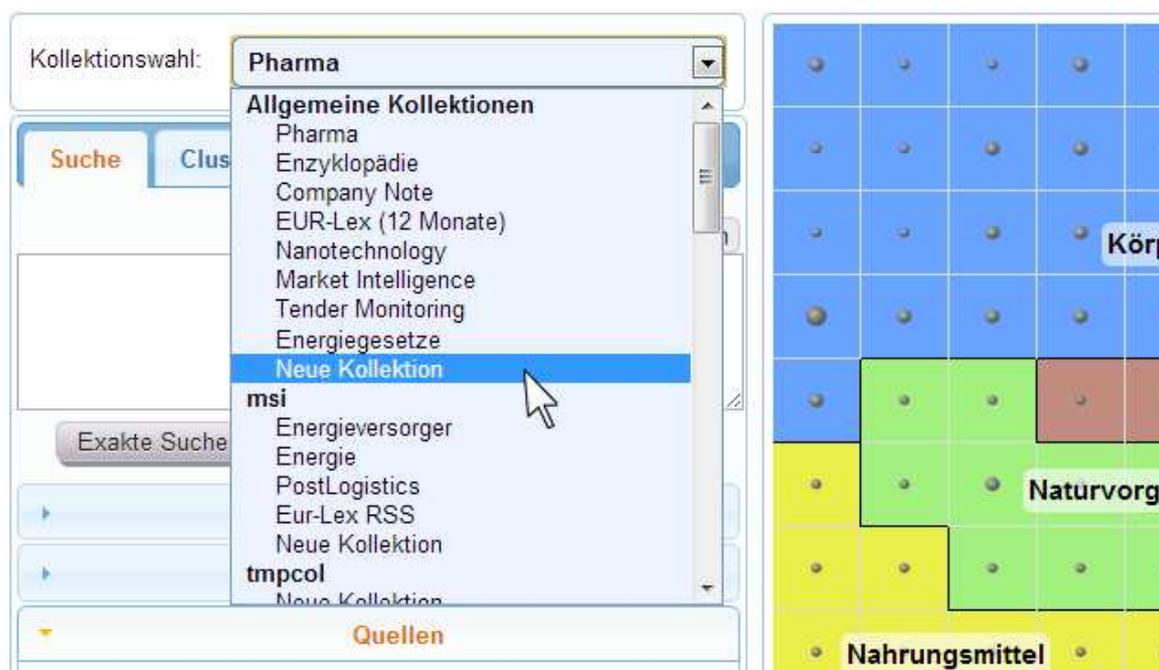


Abb. 9: Kollektionsauswahl mit den Domänen „Allgemein“ (Hauptdomäne), „msi“ und „tmpcol“. Unter jeder Domäne sind die verfügbaren Kollektionen und der Menüpunkt „Neue Kollektion“ aufgelistet

4.2 Eine Kollektion anlegen

Benutzer mit ausreichenden Rechten sehen am Ende der Kollektionsliste unter jeder Domäne einen Eintrag „Neue Kollektion“. Bei Auswahl dieses Eintrags wird ein Fenster zur Erstellung einer neuen Kollektion geöffnet.

4.3 Die Eigenschaften einer Kollektion festlegen

Bei der Erstellung werden die Bezeichnung für die neue Kollektion und ein paar Optionen abgefragt:

- Import-Modus:
Im Update-Modus wird bei Änderungen an einem Dokument der bestehende Eintrag in der Kollektion nachgetragen. Das ist der Normalfall. Wird das Häkchen entfernt, erstellt InfoCodex bei jeder Änderung einen neuen Eintrag für das Dokument.
- Abstracts generieren, Dokumentfamilien bilden:
Vgl. Kapitel 3.4 und 3.5.
- Abstandssuche, Highlighting:

Die Abstandssuche ermöglicht es, nach im Dokument nahe beieinander liegenden Begriffen zu suchen. Zusätzlich werden vereinfachte Textversionen aller importierten Dokumente zwischengespeichert, die auf Wunsch mit zur Suche passenden farblichen Hervorhebungen angezeigt werden können.

Neue Dokumenten-Kollektion eröffnen

Kollektions-Bezeichnung	<input type="text"/>
Import-Modus	<input checked="" type="checkbox"/> Update, d.h. bei späterem Hinzufügen (Incremental Load) ist bei Dokumenten mit gleichen Filenamen der alte Eintrag zu aktualisieren (andernfalls werden die Dokumente bei jedem Import zusätzlich hinzugefügt)
Abstracts generieren	<input checked="" type="checkbox"/> ja (braucht aber etwas Zeit)
Dokumentenfamilien bilden	<input checked="" type="checkbox"/> ja (ähnliche Dok. identifizieren)
Abstandssuche/Highlighting	<input checked="" type="checkbox"/> ja (aktivieren)
<input type="button" value="Erweiterte Vorgaben"/>	
<input type="button" value="OK Dokumente hinzufügen"/> <input type="button" value="OK leere Kollektion speichern"/> <input type="button" value="Abbrechen"/>	

Abb. 10: Neue Kollektion eröffnen

4.4 Erweiterte Vorgaben für die Kollektionsgestaltung

Mit den unter Abschnitt 4.3 vorgegebenen Einstellungen führt InfoCodex alle Operationen vollautomatisch aus. Wenn der Benutzer Einfluss auf die Kategorisierung, die Bildung der Deskriptoren, die Extraktion von Metadaten oder das Setzen von Zugriffsrechten nehmen will, kann mit dem Link "Erweiterte Vorgaben" die volle Maske abgerufen werden.

Nach der Erstellung der leeren Kollektion bietet InfoCodex an, Datenquellen für den Import von Dokumenten auszuwählen. Dieses Vorgehen ist in Kapitel 5 näher beschrieben.

Aktuelle Kollektion: **Pharma**

Kollektion einrichten	
Kollektions-ID:	<input type="text" value="1"/> Status: <input type="text" value="-1"/>
Kollektions-Bezeichnung	
deutsch:	<input type="text" value="Pharma"/> Muss
englisch:	<input type="text" value="Pharma"/>
französisch:	<input type="text" value="Pharma"/>
italienisch:	<input type="text" value="Pharma"/>
Kollektions-Ordner	<input type="text" value="D:\infocodex\data1"/>
User-Gruppen	<input type="text" value="1,10,13,2,20,6"/> Muss
Optionale Klassifizierungs-Vorgaben	
Klassifizierung	
Keyword-Tabelle	<input type="text"/>
Vordefinierte Kategorien	<input type="text"/>
Instruktion für Metadaten	<input type="text"/>
Vorgelagerte Datenbank	<input type="text"/>
Abstandssuche/Highlighting	<input checked="" type="checkbox"/> ja (aktivieren)
Spezielle Vorgaben	<input checked="" type="checkbox"/> Abstracts generieren <input checked="" type="checkbox"/> Dokumentenfamilien bilden
Import-Modus	<input checked="" type="checkbox"/> bei gleichen Filenamen → alten Eintrag aktualisieren
Kartengestaltung	
Anzahl Spalten der Karte	<input type="text"/> (Anz. Spalten = Anz. Zeilen)
Anzahl Hauptgebiete	<input type="text"/> (Grobunterteilung der Karte)
<input type="button" value="OK-Speichern"/> <input type="button" value="Abbrechen"/>	

Abb. 11: Erweiterte Vorgaben für eine Kollektion

User-Gruppen	Beschränkung der Zugriffsrechte auf spezielle User-Gruppen
Keyword-Tabelle	Tabelle, welche besonders wichtige Wörter/Terme und deren relative Gewichtung enthält. Diese Schlüsselwörter beeinflussen einerseits die Kategorisierung und andererseits die Bildung von Deskriptoren (siehe Benutzerhandbuch Teil 2, Abschnitt [5.1])
Vordefinierte Kategorien	Tabelle, welche die Kategorisierung steuert oder sogar fest vorgibt (siehe Benutzerhandbuch Teil 2, Abschnitt [5.3])
Instruktion für Metadaten	Mittel für die Spezifikation von Vorschriften für die Extraktion von speziellen Metadaten aus den Dokumenten
Vorgelagerte Datenbank	Der linguistischen Datenbank von InfoCodex (mit 3 Millionen Einträgen) kann eine fachspezifische individuelle Datenbank, einschliesslich Taxonomie, vorgelagert werden. Sie kann z.B. firmeninterne Abkürzungen oder spezielle Fachausdrücke enthalten. Die vorgelagerte Datenbank hat Vorrang gegenüber der linguistischen Datenbank von InfoCodex. Mit ihrer Hilfe können die Kategorisierung und die Bildung von Deskriptoren beeinflusst werden (siehe Benutzerhandbuch Teil 2, Abschnitt [5.2])
Import-Modus	Beim Hinzufügen von Dokumenten zu bestehenden Kollektionen werden die zusätzlichen Dokumente im Default-Fall auch dann in die Kollektion aufgenommen, wenn sie bereits früher importiert worden sind. Mit dem Anhaken der speziellen Option "Import-Modus" wird erreicht, dass beim Import eines Dokuments mit einem bereits registrierten Dateinamen kein zweiter Eintrag, sondern lediglich eine Aktualisierung erfolgt.

5. Dokumente hinzufügen

5.1 Wahl der Datenquellen

Für eine einzelne Kollektion können maximal 200 verschiedene Datenquellen spezifiziert werden. Es werden alle Dokumente aus den angegebenen Datenquellen in die Kollektion aufgenommen, soweit InfoCodex die gefundenen Dokumente analysieren kann.

Die Originale werden dabei nicht kopiert, sondern nur gelesen, analysiert und indiziert. Die Charakteristiken werden zusammen mit dem Speicherort des Originaldokuments in die InfoCodex-Datenbank aufgenommen. Die Ablage der Dokumente kann dabei beliebig sein und hat keinen Einfluss auf die durch InfoCodex gebildete virtuelle Ordnung.

Wahl der Datenquellen

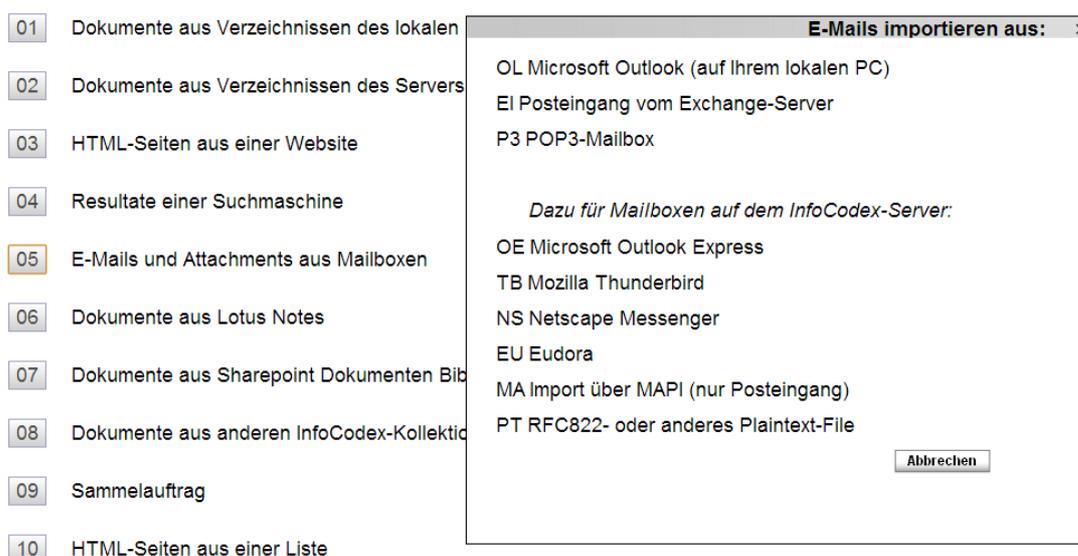


Abb. 12: Wahl der Datenquellen für den Dokumentimport

Dokumente vom eigenen Client

Die InfoCodex-Applikation und der Browser haben aus Sicherheitsgründen keinen direkten Zugriff auf die Dokumente auf dem lokalen Client. Um Dokumente vom lokalen Client in InfoCodex aufzunehmen, muss vorgängig das Plug-In auf dem lokalen PC installiert werden, das auf der Startseite von InfoCodex heruntergeladen werden kann.

Die Installation braucht möglicherweise Administratorrechte. Nach der Installation sollte man sich vergewissern, dass der Benutzer die nötigen Zugriffsrechte auf die installierten Programme hat. Beim Import vom lokalen Client wird ein Ordner in gewohnter Weise selektiert, und anschliessend werden alle Dokumente dieses Ordners in InfoCodex aufgenommen.

Dokumente auf dem Server (inkl. Netzwerk)

Dies ist der häufigste Fall, um Dokumente aus dem internen Netzwerk zu importieren. Es können alle Laufwerke und Netzwerkpfade angewählt werden, auf die der Server Zugriff hat und für die der Benutzer die erforderlichen Zugriffsrechte besitzt.

Websites

Es braucht nur die Startadresse der gewünschten Website angegeben zu werden, z.B. . Infocodex verfolgt alle Links bis einer spezifizierten Tiefe und lädt die gefundenen Seiten zur Inhaltsanalyse herunter. Es können alle von InfoCodex unterstützten Dateitypen importiert werden; bei Bedarf kann jede Datenquelle auf bestimmte Dateitypen beschränkt werden.

Für passwortgeschützte Websites kann im Bedarfsfall ein Passwort mitgegeben werden.

Resultate einer Suchmaschine

Nach der Wahl einer Suchmaschine (z.B. Google) erscheint eine Maske, in der man die Suchabfrage wie gewohnt eingibt. InfoCodex leitet die Anfrage an die ausgewählte Suchmaschine weiter und lädt alle gefundenen Dokumente zur Inhaltsanalyse herunter. Dabei ist zu beachten, dass die meisten Suchmaschinen höchstens 1'000 Dokumente liefern, und zwar auch dann, wenn die angezeigte Trefferzahl weit darüber liegt.

Die Liste der verfügbaren Suchmaschinen kann je nach Installation variieren.

Für passwortgeschützte Suchmaschinen (z.B. Web of Science) kann im Bedarfsfall ein Passwort direkt eingegeben oder vom Systemadministrator hinterlegt werden.

E-Mails und Attachments

InfoCodex unterstützt verschiedene Mailprogramme wie Outlook, Thunderbird, Outlook Express, Netscape, Eudora u.a.m. Der Inhalt von Attachments wird mit einbezogen. Ein E-Mail wird zusammen mit seinen Attachments als ein Dokument betrachtet.

Für die Wahl

Microsoft Outlook (auf Ihrem lokalen PC)

bzw.

Posteingang vom Exchange-Server

muss Outlook auf dem InfoCodex-Server installiert sein. Im zweiten Fall muss InfoCodex auch Zugriff auf den Exchange-Server haben, und der Benutzer muss seinen Posteingang für InfoCodex freigeben.

Der Import von Outlook auf dem lokalen PC kann nicht vollautomatisch erfolgen, weil der Benutzer aus Sicherheitsgründen alle paar Minuten den Zugriff auf seine Nachrichten bestätigen muss. Es ist in der Regel sinnvoller, die gespeicherten Mails als Outlook-Archiv (PST- File) auf den Server zu stellen und von dort aus als Datei zu importieren. InfoCodex erkennt solche PST-Archive und kann die enthaltenen Elemente automatisch extrahieren.

Dokumente aus Lotus Notes

Für den Zugriff auf Lotus Notes müssen ein Notes-Client und das InfoCodex-Modul Notes Connect auf dem InfoCodex-Server installiert werden.

Beim Import werden die Lotus-Notes-Dokumente vorgängig mit Notes Connect aus Lotus Notes heraus selektiert und als Index-Dateien im Notes Connect-Verzeichnis für die Verarbeitung durch InfoCodex zwischengespeichert.

Dokumente aus SharePoint

Dokumente aus SharePoint können ganz einfach über die URL einer Dokumentenbibliothek importiert werden. Zusätzlich können bei Bedarf Benutzername und Passwort angegeben werden.

Dokumente aus anderen Kollektionen

Im Bedarfsfall können auch Dokumente aus bestehenden InfoCodex-Kollektionen übernommen werden, und zwar als Resultat einer Suchabfrage (analog zu den Resultaten von Web-Suchmaschinen).

Sammelauftrag

Im Sammelauftrag können mehrere Anfragen an verschiedene Suchmaschinen gleichzeitig erfasst und aufgegeben und zusätzlich mit Dokumenten aus Websites und lokalen Ordnern ergänzt werden. Dieses Mittel dient der Rationalisierung von Recherchen.

HTML-Seiten aus einer Liste

Diese Funktion ermöglicht den Import von Dokumenten aus dem Internet, die in einer Listendatei bereit stehen.

5.2 Importe ausführen (Inhaltsanalyse/Indexierung)

Nach der Auswahl der Datenquellen, die in die Kollektion einbezogen werden sollen, kann die automatische Verarbeitung mit "**OK Importieren**" ausgelöst werden. Vorgängig können mit **Optionen** noch allfällige Sonderbedingungen gesetzt werden, z.B. ein Maximalalter der aufzunehmenden Dokumente oder die Aktivierung eines OCR-Interpreters für gescannte Dokumente.

- 09 Sammelauftrag
- 10 HTML-Seiten aus einer Liste



Es laufen nun die folgenden Prozesse ab:

1. Die Spider-Agents von InfoCodex sammeln die Dokumente, konvertieren diese temporär in Textdateien und extrahieren die Metadaten aus den Dokumenten (Autor, Titel, Datum etc.)
2. Spracherkennung: deutsch / englisch / französisch / italienisch / spanisch; lexikalische und semantische Analyse
3. Sprachübergreifende Inhaltsanalyse, Verknüpfung mit der Taxonomie
4. Konstruktion eines 100-dimensionalen, auf die Kollektion ausgerichteten Inhaltsraums und Projektion der Dokumente in diesen Raum
5. Kategorisierung der Dokumente mittels selbstorganisierender Neuronaler Netzwerk-Technik (Kohonen-Map) ↓ Einordnung in ein Büchergestell
6. Verschlagwortung der Dokumente: Bildung von Deskriptoren
7. Indexierung für die effiziente Suche.

Diese Prozesse laufen im Hintergrund ab. Der Benutzer braucht diese Prozesse nicht zu verfolgen und kann sich anderen Aufgaben zuwenden.

Job-Scheduling

Anstelle einer direkten Ausführung können die Import-Anweisungen auch mit **Job-Scheduling** als Batch-Job aufgegeben werden. Startzeitpunkt und Periodizität des auszuführenden Imports werden im Dialog abgefragt. Die Batch-Anweisungen können in InfoCodex jederzeit nachgeschaut und im Bedarfsfall editiert oder gelöscht werden:

→ Admin → C2 Datenquellen sichten/ändern.

Typische Anwendungen

Konkurrenzüberwachung: Jede Nacht werden die neuesten, über die Tätigkeiten der Konkurrenz verfügbaren Informationen gesammelt. Diese Informationen stehen am Morgen in geordneter und übersichtlicher Form zur Verfügung.

Aktualisierung grosser Kollektionen: Die Dokumente des lokalen Netzwerkes werden über das Wochenende neu geladen ("Initial Load"). Montag bis Donnerstag werden die neuesten Dokumente, die höchstens einen Tag alt sind, jeweils über Nacht hinzugefügt ("Incremental Load").

Importoptionen

Die folgenden zusätzlichen Importoptionen sind verfügbar:

Kollektion: Pharma

Importe in Warteliste

1. Website "http://www.widas.com/" kopieren (bis zu 1000 Dok. von www.widas.com)

Maximalalter der Dokumente in Tagen, z.B. 30 oder 30+
(**30+ heisst:** inkl. Dokumente mit unbekanntem Datum)

- Dokumente als TXT-Files zwischenspeichern
- nicht konvertierbare Files merken
- wende Content-Filter auf HTML Seiten an

Min. anzahl Zeichen pro Textblock für Content Filter

Berechtigungen vergeben/ OCR aktivieren

Import Job

berechtigte Usergruppen

(:T:/20) Anzahl Jobs

OCR

OK

Abbrechen

Maximalalter Maximalalter der Dokumente in Tagen. Dadurch werden nur neuere Dokumente importiert, was speziell bei periodischen "Incremental-Loads" grosser Kollektionen die Prozesszeit verkürzt.

TXT-Files Dokumente als TXT-Files zwischenspeichern.

Nicht konvertierbare Files merken Nicht konvertierbare Files merkt sich das System auf einer Blacklist und verkürzt dadurch bei periodischen Incremental Loads grosser Kollektionen die Prozesszeit.

Content-Filter HTML-Seiten beinhalten nebst dem eigentlichen Content meist zusätzliche Informationen wie Navigation und Werbung. Der Content-Filter macht sich zunutze, dass Content-Textblöcke meist grösser sind als die Textblöcke der Zusatzinformationen.

Ein Setzen des Content-Filters pro Datenquelle ist möglich durch Eintragen einer CONFILT-Zeile (z.B. CONFILT=60) im Instruktions-File des Batch-Jobs.

Berechtigungen	Bei Import-Jobs bestehend aus mehreren Datenquellen können für jede Quelle einzeln die Berechtigungen vergeben werden.
OCR	OCR-Texterkennung für grafische Dokumente einschalten. Standardmässig benutzt InfoCodex die OCR-Engine von Microsoft, welche Teil von Microsoft Office (ab Version 2003) ist. Unterstützung anderer OCR-Engines (OmniPage, FineReader) ist optional möglich.

6. Inhalte auswerten

Die Auswertungsfunktionen können über das Menü **Inhalt** erreicht werden.

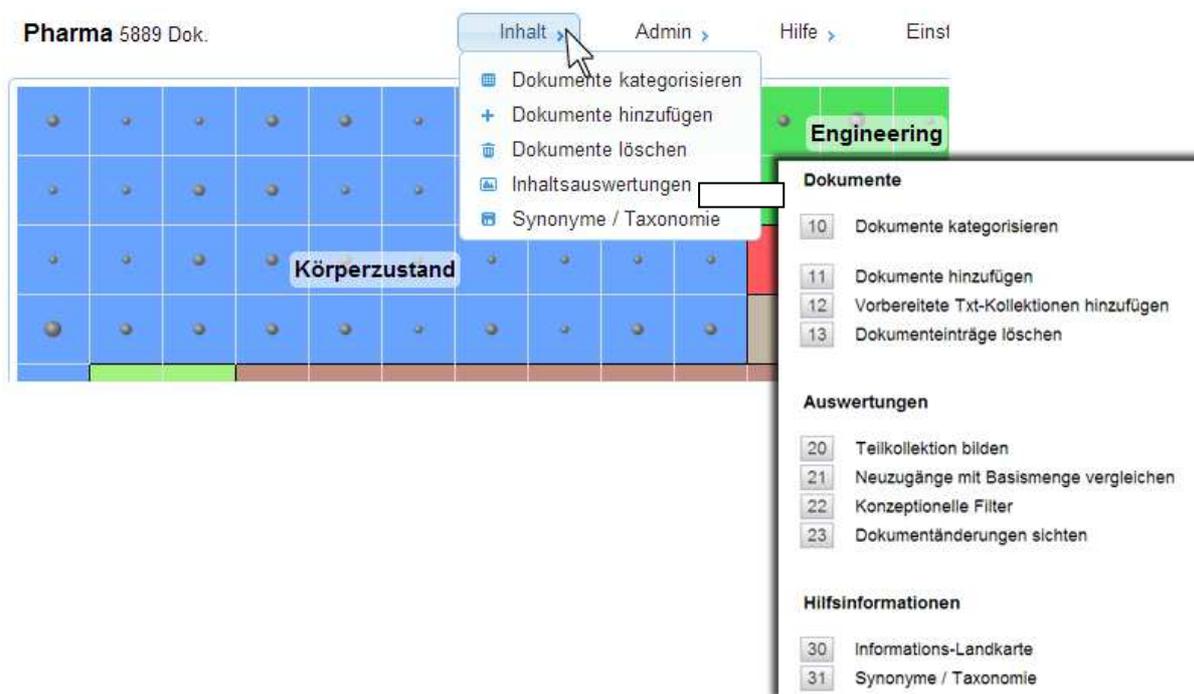


Abb. 13: Inhalte: Die eigentlichen Analysen stehen unter **Inhaltsauswertungen** zur Verfügung.

6.1 Dokumente kategorisieren

Mit dieser Funktion werden die Dokumente von wählbaren Datenquellen aufgrund ihres Inhalts den einzelnen Fächern der Informations-Landkarte zugeordnet. Die Dokumente innerhalb eines Faches haben ähnliche Inhalte, und damit bedeutet eine solche Zuordnung eine *Kategorisierung nach inhaltlichen Gesichtspunkten*.

Diese Funktion ist auch eine Grundlage für ein Response-Management: Eingehende Dokumente werden kategorisiert, und aufgrund der zugewiesenen Kategorie werden mit Hilfe einer bereitgestellten Entscheidungsmatrix gewisse Folgeaktionen ausgelöst.

Anwendungsbeispiele:

- Automatische Verteilung von E-Mails, die an eine zentrale Adresse eingehen.
- Automatische Vorverarbeitung von Kundenanfragen/Kundenreklamationen mit Vorbereitung einer Antwort aufgrund bereitgestellter Textbausteine
- Suchen von Normen und Vorschriften zu bestimmten Problemstellungen.

Neben der Kategorisierung findet auch eine automatische Verschlagwortung der Dokumente mit 18 Deskriptoren statt. Diese Deskriptoren, welche in der Regel aussagekräftig sind, können für die Verschlagwortung von Dokumenten bei der Einlagerung in Archive verwendet werden.

Ablauf:

1. Aufruf der Funktion "Dokumente kategorisieren"
2. Wahl der Datenquellen (genau gleich wie in Abschnitt 4.2 beschrieben)
3. Ausgabe der Resultate auf dem Bildschirm oder in einer Exceltabelle (als wählbare Zusammenstellung von Dokument-Merkmalen und Deskriptoren)

Bei dieser Funktion werden die analysierten Dokumente nicht in die bestehende Kollektion aufgenommen.

6.2 Dokumente hinzufügen

Mit diesem Menüpunkt können neue Datenquellen zur aktuellen Kollektion hinzugefügt werden; vgl. Kapitel 5.1.

6.3 Dokumenteinträge löschen

Mit dieser Funktion können Dokumente aus der aktuellen Kollektion entfernt werden. Dabei werden nur die Verweise auf das Dokument in der Kollektion gelöscht; die Original-Dokumente bleiben unangetastet.

Vorgehen:

1. Suche aufgeben, welche die zu löschenden Einträge enthält.
2. Speichern dieser Suche.
3. Nach dem Aufrufen dieser Suche werden die effektiv zu löschenden Einträge in einer Feinselektion festgehalten.
4. Anschliessend erfolgt das Löschen.

6.4 Teilkollektionen bilden (unter „Inhaltsauswertungen“)

Diese Funktion bildet eine neue Kollektion als Teilmenge der aktuellen Kollektion. Die Teilmenge wird durch eine Selektion von ausgewählten Hauptthemen und/oder den Resultaten einer gespeicherten Suche gebildet.

6.5 Neuzugänge mit der Basiskollektion vergleichen / Trenderkennung

Diese Funktion bietet Früherkennung von Veränderungen, z.B. bei:

- Konkurrenzüberwachung/ Marktbeobachtung
- Nachrichtendienst
- Erkennen von neuen Trends und Technologien

Typischerweise wird eine Kollektion zu einem bestimmten Thema angelegt (z.B. eine Kombination mehrerer Internetrecherchen). In periodischen Abständen werden die neuesten Dokumente zu dieser Kollektion hinzugefügt.

Durch Vergleich der Neuzugänge mit der Basis-Kollektion sollen die wichtigsten Veränderungen rasch und zuverlässig erkannt werden.

Mit den drei in der Maske beschriebenen Auswertungen: „Thematische Verschiebungen“, „Heatmap“, „Neue Themen“ können die Veränderungen auf Knopfdruck sichtbar gemacht werden. Unter „Neue Themen“ erscheinen beispielsweise diejenigen Dokumente, welche gegenüber den bisherigen Dokumenten echt neue Fakten enthalten.

Neuzugänge analysieren

Zurück Suche Inhalte Admin ? Exit

CentreDoc-Website : Neuzugänge analysieren

Es werden die inhaltlichen Schwergewichte von Neuzugängen mit denjenigen einer Basismenge verglichen.

Selektion Basismenge und Neuzugänge

Importdatum	Anzahl Dokumente	Basismenge	Neuzugang
23.03.06	584	<input checked="" type="checkbox"/>	<input type="checkbox"/>
24.03.06	100	<input type="checkbox"/>	<input type="checkbox"/>
3.04.06	30	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Mögliche Auswertungen

Thematische Verschiebungen	Auflisten der wesentlichen inhaltlichen Verschiebungen
Heat-Map	Die Felder (Neuronen) mit dem grössten relativen Zuwachs werden rot gefärbt
Neue Themen	Neuzugänge sortiert nach grössten Abweichungen vom Bisherigen

Thematische Verschiebungen Heat-Map Neue Themen Abbrechen

Abb. 14: Neuzugänge analysieren; thematische Verschiebungen sichtbar machen

7. Synonyme / Taxonomie

Diese Funktion gibt einen Einblick in die linguistische Infrastruktur von InfoCodex. Sie dient insbesondere auch dazu, den Umfang von Synonym-Gruppen zu sichten und in Zweifelsfällen eine Klärung herbeizuführen (nämlich dann, wenn einem Dokument ein seltsam erscheinender Deskriptor zugewiesen wird, den man bei einer ersten Betrachtung nicht mit dem Dokument in Verbindung bringen kann).

Hinweis: Die einzelnen Wörter einer Sprache haben häufig eine mehrfache Bedeutung, und die Übersetzung in andere Sprachen bzw. die Einordnung in eine Synonymgruppe und die Zuweisung in die Taxonomie sind dementsprechend subjektiv. Für die Gesamtklassifikation eines Dokumentes durch das Neuronale Netz ist dies im Allgemeinen nicht problematisch. Solche Unschärfen liegen auch beim Lesen und Aufnehmen von Dokumentinhalten durch den Menschen vor.

Synonyme → ähnliche Wörter anzeigen

Wort eingeben, dessen Synonyme in 3 wählbaren Sprachen anzuzeigen sind

Ausgabe in englisch deutsch französisch italienisch spanisch

Wort deutsch
oder englisch

OK

oder

Begriffshierarchie → Klassifikationssystem (Taxonomie)

Begriff eingeben, dessen Einordnung Sie sehen wollen (kein Begriff = Übersicht)

Begriff deutsch
oder englisch

OK

Abb. 15: Abfragen von Synonymen und Begriffshierarchien

7.1 Synonymgruppen

Mit dem oberen Teil der Maske (**Synonyme**) kann nachgeschaut werden, welcher Synonymgruppe ein bestimmtes Wort zugeteilt wird und wie dieses Wort in der Begriffshierarchie eingestuft ist. Die Begriffe werden in drei wählbaren Sprachen nebeneinander angezeigt.

Eine Synonymgruppe fasst mehrere Begriffe mit gleicher Bedeutung zusammen, z.B. „Fahrausweis“, „Führerschein“, „driving licence“, „permis de conduire“ usw.

Internet >> Internet >> Computer-Netzwerk >> Informatik >> INFORMATION/ KOMMUNIKATION

deutsch

Internet, Internet 2, Internetz
Internetwelt, Internet der
*internetadäquat, internetbasierende
internetbasiert, internetbezogen
internetbezogene, internetgestützt
internetgestützten, internettauglich*

englisch

internet, internet world, internets
Internet of

französisch

Internet, Internet des
internet, internet haut
l'internet

Abb. 16: Mit dem Wort „Internet“ verwandte Begriffe in Deutsch, Englisch und Französisch

7.2 Begriffshierarchie

Mit dem unteren Teil der Maske (**Begriffshierarchie**) erhält der Benutzer einen Einblick in das gesamte Klassifikationssystem („Taxonomie“), und zwar von ganz oben bis hinunter zu den einzelnen Synonymgruppen.

Wenn in den Feldern der Begriffshierarchie keine Einschränkung eingegeben wird, dann zeigt InfoCodex eine Übersicht über die beiden obersten Stufen des Taxonomiebaums.

Bei Vorliegen von individuellen vorgelagerten Datenbanken werden diese Strukturen ebenfalls berücksichtigt. Für den Benutzer ist dabei ersichtlich, ob die Begriffe zur vorgelagerten Datenbank bzw. zur Standard-Datenbank von InfoCodex gehören.

Die Verknüpfung der Synonymgruppen mit dem Taxonomiebaum, d.h. die Zuordnung von Hypernymen (Oberbegriffen) zu den Synonymgruppen, entspricht einer speziellen Ontologie, in welcher man sich bewusst auf die wichtigste Bedeutung der Wörter bezieht.

Informatik >> INFORMATION/ KOMMUNIKATION

0	• Informatik	• information technology
1	• Software	• software
1.01	• Business-Software	• business software
1.0101	• Adressbewirtschaftung	• address management
1.0102	• CAD	• computer-aided design
1.0103	• Marketing-Datenbank	• database marketing
1.0104	• Finanzsoftware	• financial software
1.0105	• HR software	• HR software
1.0106	• Wissens-Management	• knowledge management
1.0107	• Suchtechnologie	• search technology
1.0108	• Auftragsbewirtschaftung	• order processing
6.01	• Kommunikations-Software	• communication software
6.02	• Identifikationskarte	• identification cards
6.03	• Computervisualisierung	• computer graphics
6.04	• Bilddaten	• image data
7	• Computer-Netzwerk	• computer network
7.01	• WLAN	• WLAN
7.02	• Internet	• internet
7.03	• Intranet	• intranet
8	• Peripheriegerät	• peripheral equipment
8.01	• Datenspeicherungsgerät	• data storage device
8.0101	• Disk	• disk
8.02	• Verbindungsausrüstung	• interconnection equipment

Abb. 17: Begriffshierarchie für das Gebiet „Informatik/Kommunikation“, dem der Begriff „Internet“ zugeordnet ist.

8. System-Management (Kurzfassung von Teil 2 des Handbuchs)

Die System-Management-Funktionen von InfoCodex werden über den Menüpunkt

 System-Administration aufgerufen. Sie sind gegliedert nach:

User-Administration	Zugelassene Benutzer und Benutzergruppen Datenschutzaspekte LDAP-Schnittstelle zur zentralen Benutzerverwaltung Einrichten von autonomen InfoCodex--Domänen
System-Administration	Verwaltung der Kollektionen Beeinflussung der Kategorisierung Systemeinstellungen

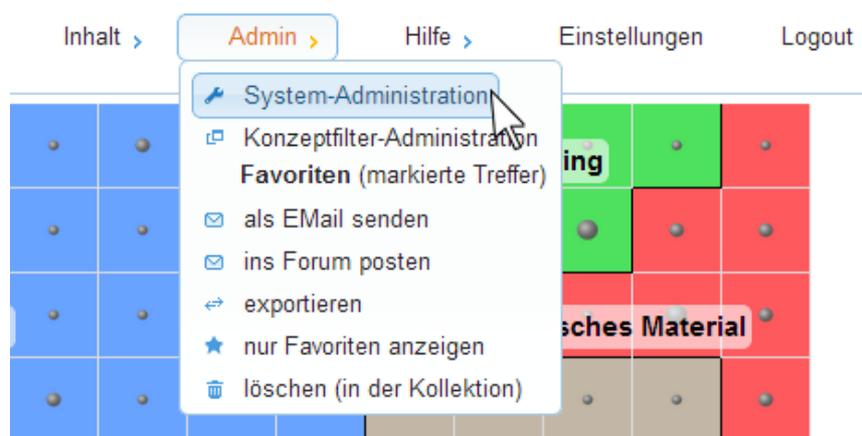


Abb. 18: Menü „Admin“

Diese Funktionen sind in Teil 2 des Benutzerhandbuchs beschrieben. In diesem Kapitel werden lediglich einige Aspekte, die auch für den Anwender von Bedeutung sind, kurz zusammengefasst.

Für normale Benutzer sind diese Funktionen nur zum Teil sichtbar. Unter "User-Administration" kann beispielsweise nur die Funktion "Passwort ändern" erscheinen.

8.1 Datenschutz / Domänen-Konzept

Der Datenschutz wird in InfoCodex auf drei verschiedenen Ebenen gewährleistet:

Benutzergruppen

Jeder InfoCodex-Benutzer ist Mitglied einer oder mehrerer Benutzergruppen. Zugriffsrechte auf Domänen und Kollektionen können pro Benutzer und Gruppe gesetzt werden. Die effektiven Rechte eines Benutzers ergeben sich durch seine Rechte und der Rechte aller Gruppen, denen er angehört.

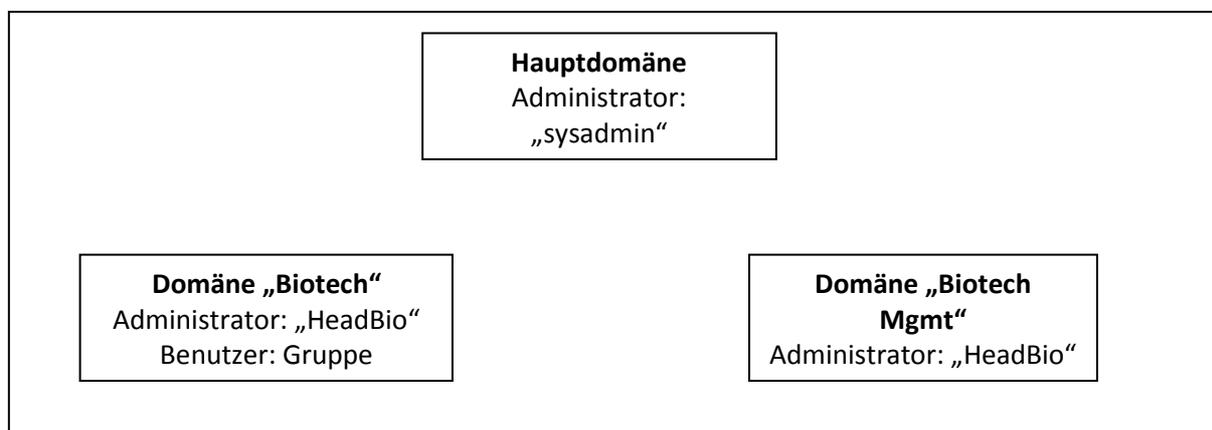
File System Security

Wenn diese Option aktiviert wurde, kann ein Benutzer nur Dokumente aus Datenquellen finden und sichten, auf die ihm im Netzwerk auch ausserhalb von InfoCodex ausreichende Zugriffsrechte gewährt wurden. Alle Zugriffe auf Dateien erfolgen dann im Kontext des jeweiligen Benutzers, so dass die Zugriffskontrolle und Protokollierung auf Betriebssystemebene gewährleistet sind.

InfoCodex-Domänen

Der Datenbereich von InfoCodex, in welchem die Kollektionen abgelegt sind, kann in eine Hauptdomäne und eine beliebige Anzahl von Subdomänen mit individuellen Administratorrechten unterteilt werden.

Zugriff auf eine Kollektion bedingt entsprechende Zugriffsrechte auf die Domäne, in der die Kollektion liegt. So kann z.B. eine geschützte Domäne für Mitglieder der Geschäftsleitung eingerichtet werden, auf die selbst die IT-Abteilung keinen Zugriff hat.



Diese Funktionen beziehen sich auf das Einrichten und die Pflege der einzelnen Kollektionen (vgl. Abschnitt 4). Sie sind nur denjenigen Benutzern zugänglich, die über die entsprechenden Rechte verfügen.

Spezielle Funktionen:

Kollektion löschen	Die gesamte Kollektion wird aus dem System entfernt.
Status sichten	Den Verarbeitungsstatus einer Kollektion sichten (speziell in der Import- und Analyse-Phase); Verarbeitungsprotokoll sichten
Job-Scheduling	Falls der Import von Dokumenten durch Batch-Jobs gesteuert ist (vgl. Abschnitt 5.2), können die entsprechenden Scripts unter dem Menüpunkt "Datenquellen sichten/ändern" gesichtet und allenfalls gelöscht oder editiert werden.

8.3 Beeinflussen der Kategorisierung

Die Kategorisierung der Dokumente (Einordnung in ein sachlogisch aufgebautes Büchergestell) und die Vergabe von Deskriptoren für die einzelnen Dokumente wird massgeblich durch die linguistische Datenbank und die Taxonomie von InfoCodex bestimmt. Diese Datenbank enthält mehr als drei Millionen Ausdrücke in englischer, deutscher, französischer, italienischer und spanischer Sprache, und sie deckt im Wesentlichen alle Wissensgebiete ab.

Es kann dennoch wünschenswert sein, dass der Anwender seine eigenen Akzente setzen und beispielsweise spezielle Fachausdrücke oder firmenspezifische Abkürzungen einführen kann. Damit können die Kategorisierung und das automatische Generieren der Deskriptoren durch den Anwender wesentlich beeinflusst werden.

Folgende Mittel stehen zur Verfügung:

Keyword-Setting

Es kann eine Liste von Wörtern/Ausdrücken definiert werden, die bei der Inhaltsanalyse besonders stark zu gewichten sind.

Vorgelagerte Datenbank

Es kann eine kundenspezifische linguistische Datenbank (mit Links zu der Taxonomie) bereit gestellt werden, welche spezielle Fachausdrücke bzw. firmeninterne Abkürzungen enthält. Eine solche vorgelagerte Datenbank hat Priorität gegenüber der Standard-Datenbank von InfoCodex. Wenn bei der Inhaltsanalyse ein Wort in der vorgelagerten Datenbank gefunden wird, dann ist die entsprechende Sinndeutung massgebend; andernfalls kommt die InfoCodex-Datenbank zum Tragen.

Vordefinierte Kategorien

Mit diesem Mittel kann die Gliederung der Informations-Landkarte (d.h. des "Büchergestells") beeinflusst werden. In vielen Fällen kann mit der interaktiven und einfachen "Ad-hoc-Kategorisierung" viel erreicht werden. Weiter besteht auch die Möglichkeit, fixe Kategorien vorzugeben und den Automatismus des selbstorganisierenden Neuronalen Netzes auszuschalten. So kann z.B. eine bestehende firmeninterne Klassifizierung in InfoCodex abgebildet werden.

9. Ausgewählte Beispiele

9.1 Informations-Recherchen

Problemstellung

Es soll periodisch nach Erkenntnissen zum Thema "Sekundäre Vergiftungen von Greifvögeln nach dem Einsatz von Giftködern gegen *Arvicola terrestris*" gesucht werden.

Suchbegriffe: Rodentizide, Bromadiolone, Greifvögel, *Arvicola terrestris*

Vorgehen

1. Eine neue Kollektion eröffnen.
2. Dokumente hinzufügen: Menüpunkt „Resultate einer Suchmaschine“, z.B. Google
3. Suchbegriffe eingeben: Rodentizid, Bromadiolon, Greifvögel, "*Arvicola terrestris*"
4. Auf „Kontrollanzeige“ klicken. Die Google-Suche ergibt 0 Treffer.
5. Auf „Erweiterte Abfragen“ klicken und die passenden Synonyme zu den einzelnen Suchbegriffen auswählen (ggf. auch in anderen Sprachen).
6. Den Import starten.

Mit diesem Vorgehen werden mehrere Google-Suchanfragen mit unterschiedlichen Synonymkombinationen generiert.

Verfeinerungen

Die erste Suche wie oben beschrieben durchführen. Danach einen Batch-Job für die gleiche Abfrage erstellen:

- wöchentlich ausführen,
- Option "2 = Dokumente zur bestehenden Kollektion hinzufügen" verwenden.

9.2 Finden von vorhandenem Know-how

Problemstellung

Ein Forscher will ein neues Forschungsgebiet erschliessen. Bei einer Recherche ist er auf ein interessantes Dokument (z.B. ein Paper) gestossen. Er will abklären, was in der eigenen Forschungsanstalt bereits für Wissen auf dem entsprechenden Gebiet vorhanden ist.

Vorgehen

1. Den gefundenen Text (Worddokument, PDF, HTML oder auch E-Mail) in die Zwischenablage kopieren.
2. In InfoCodex die Kollektion "Internes Netzwerk" wählen.
3. Den kopierten Text in das Suchfeld einfügen.
4. Die Suche starten.

Resultat

InfoCodex liefert die im eigenen Netzwerk vorhandenen Dokumente, die dem gefundenen Text möglichst ähnlich sind. Die verwendeten Begriffe brauchen dabei in den ähnlichen Dokumenten nicht enthalten zu sein; es wird lediglich nach inhaltlicher Ähnlichkeit gesucht.