Benutzerhandbuch
Teil 2: Systemmanagement

Inhaltsverzeichnis

1. Zweck des Handbuchs und Zusammenfassung

Der Teil 2 des Benutzerhandbuchs richtet sich an Systemadministratoren und privilegierte Benutzer, welche zumindest in Subdomänen über Rechte zur Benutzer- und zur Systemadministration verfügen. Es werden folgende Themen behandelt:

Systemarchitektur und Datenschutzkonzept

• *Allgemeines*	Aufbau der Standalone- und der Komponenten-Version; Diskorganisation (Software und linguistische Datenbank, Benutzerdatenbanken, Webinterface); Sicherstellungsempfehlungen
• Datenschutzkonzept	Mittel zur Gewährleistung des Zugriffsschutzes auch bei hochsensiblen Daten; Domänen-Konzept (autonome Bereiche); File-System-Security (FS-Security)

Benutzeradministration

• Benutzer und Gruppen	Benutzer und Gruppen anlegen und löschen, Rechte vergeben; Gruppenmitgliedschaften, Benutzer in eine Gruppe aufnehmen
• Zentrale Benutzerverwaltung/ LDAP-Schnittstelle	Koordination von Benutzern und Gruppen mit einer zentralen Benutzerverwaltung; ADS-Anbindung für den Anschluss an Windows-Domänen (SPoA, SSO)
• InfoCodex-Domänen	Einrichten von Autonomie-Bereichen für ausgewählte Benutzer und Gruppen; Zugriffsrechte für Domänen festlegen

Systemadministration

• *Kollektionsverwaltung*	Kollektionen einrichten und löschen; Datenquellen auswählen und konfigurieren; Reorganisation, Neugenerierung und Batch-Importe; Verarbeitungsstatus und -protokoll
• *Einflussmöglichkeiten*	Mittel zur Beeinflussung der Kategorisierung und der Verschlagwortung;
• *Import und Export von Daten*	Einlesen von benutzerspezifischen Daten ab Excel-Tabellen; Ausgabe von InfoCodex-Daten auf Excel-Tabellen

Systemeinstellungen

• *Verarbeitungssteuerung*	Setzen von Verarbeitungsoptionen; Kundenspezifische Umgebungen (Netzlaufwerke usw.)
• *Spezielle Schnittstellen*	Microsoft Exchange; Lotus Notes; Schnittstellen zu speziellen Datenbanken

2. Systemarchitektur und Datenschutzkonzept

2.1 Systemarchitektur

Wegen seiner Offenheit und Skalierbarkeit kann InfoCodex relativ leicht in bestehende Umgebungen integriert werden. Die InfoCodex-Software steht als Standalone- sowie auch als Komponenten-Version für die Einbettung in andere Systeme zur Verfügung.

Standalone-Version von InfoCodex

Diese ist Webbasiert und wird zusammen mit einem Webserver (Apache oder IIS) auf einem Server installiert. Für die Clients wird keine besondere Software benötigt; ein aktueller Webbrowser genügt.

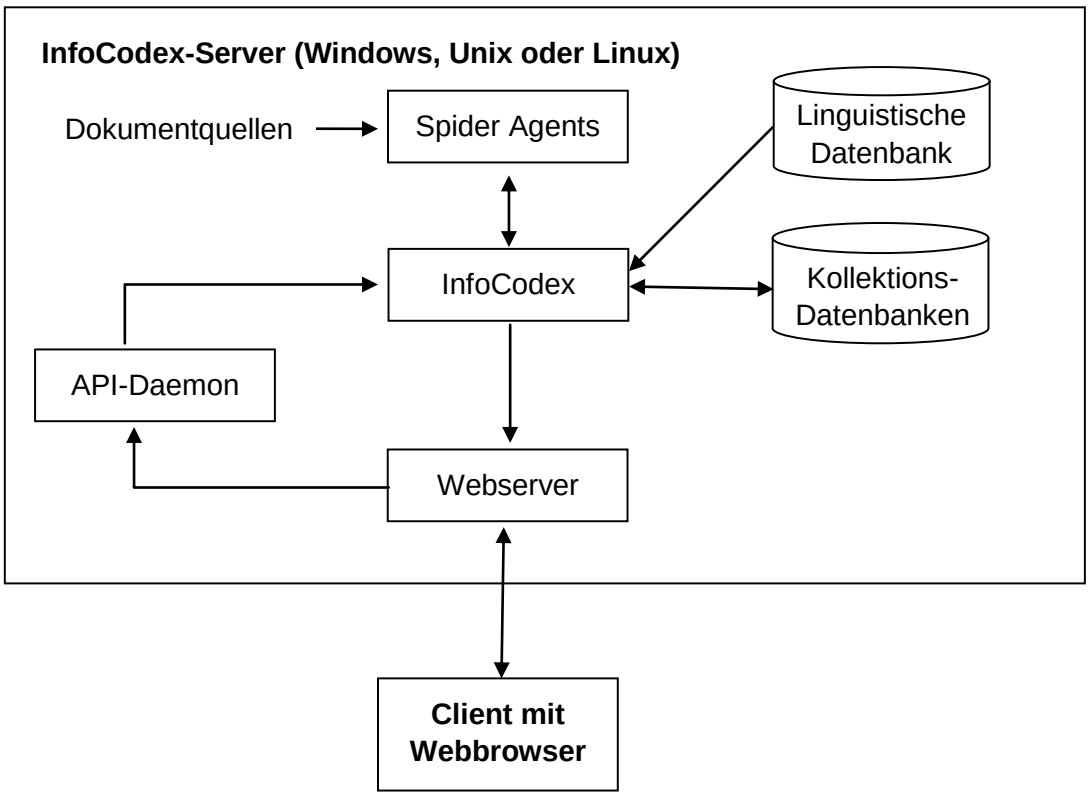


Abb. 1: Architektur der Standalone-Version

Komponenten-Version von InfoCodex

InfoCodex kann als einzelne Komponente in andere Applikationen eingebettet werden. Dank seines generischen Aufbaus sind kundenspezifische Lösungen problemlos realisierbar. Zurzeit existieren Anbindungen für C, PHP, Perl und Java. Das XML-basierte InfoCodex-API wird in Teil 3 des Benutzerhandbuchs beschrieben.

2.2 Diskorganisation

Das InfoCodex-System wird in den folgenden drei Diskbereichen installiert:

Programmbereich (stationär, erfordert keine Sicherstellung)

Dieser Bereich enthält die Software, die linguistische Datenbank, die globale Konfiguration und einige temporäre Dateien für die Prozesskoordination.

Die optionalen kundenspezifischen vorgelagerten linguistischen Datenbanken befinden sich ebenfalls in diesem Bereich. Diese werden von Exceltabellen (= Stammdaten) importiert. Sie

sind nach einem Update der linguistischen Datenbank von InfoCodex zwecks Abgleich jeweils neu zu importieren.

Datenbereich (User- und Kollektionsdaten)

Er enthält die Benutzerdatenbanken, die Kollektionsverwaltung und die einzelnen Kollektionsdatenbanken.

Es wird empfohlen, den Datenbereich in die regulären Sicherstellungsprozeduren einzubeziehen.

Web-Schnittstelle (erfordert keine Sicherstellung)

Dieser Bereich wird für die Kommunikation mit dem Browser der Benutzer benötigt (HTML-Seiten usw.).

Zugriffsrechte für die entsprechenden Verzeichnisse und Dateien

Programmbereich	**wP**
InfoCodex-Software	xP
Linguistische Datenbank	rP
Vorgelagerte Datenbanken	wP
Prozesskordination	wP

Datenbereich	**wP**
Benutzerverwaltung	wP
Kollektionsdatenbanken	wP

Web-Schnittstelle	**wP, r+**
htdocs/icd	wP, r+
htdocs/ice	wP, r+
htdocs/icf	wP, r+
htdocs/ici	wP, r+

r Lesen
w Lesen und schreiben
x Lesen und ausführen

P InfoCodex-Prozesse
\+ Alle ("world")

Falls kundenspezifische vorgelagerte Datenbanken verwendet werden, muss InfoCodex auch auf die linguistische Datenbank Schreibberechtigung haben (wP).

2.3 Datenschutzkonzept

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 werden pro Benutzer und Gruppe gesetzt. Die effektiven Rechte eines Benutzers ergeben sich aus seinen Benutzerrechten und den Rechten 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.

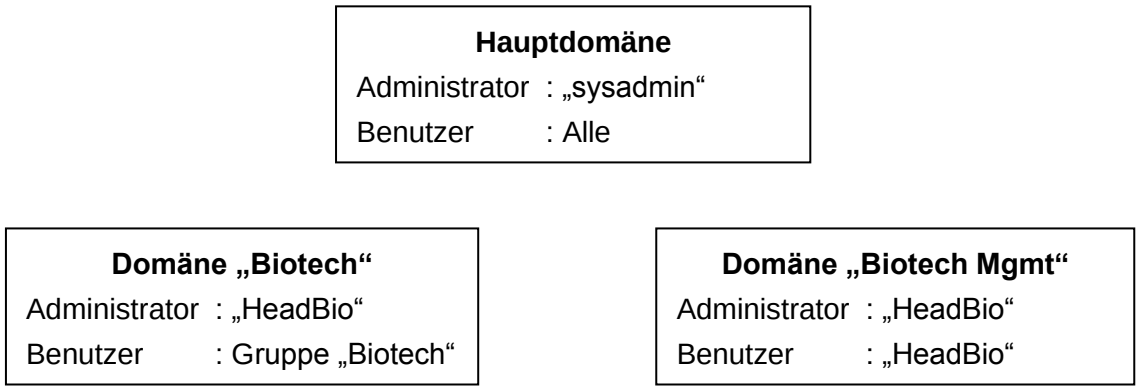


Abb. 2: InfoCodex-Domänen mit verschiedenen Benutzergruppen

Eine neue Domäne kann ausschliesslich durch den Systemadministrator der Hauptdomäne eröffnet werden. Bei der Eröffnung einer neuen Domäne wird ein Domänenadministrator ernannt, der fortan die volle Souveränität über diese IC-Domäne besitzt. Auch der Systemadministrator der Hauptdomäne hat keinen Zugriff auf die eröffnete Domäne, wenn der Domänenadministrator diesen nicht explizit freigibt. Der Systemadministrator besitzt im Normalfall lediglich das Recht, die Domäne wieder zu löschen.

Dieses Konzept erfüllt die im Umgang mit vertraulichen Informationen gestellten Sicherheitsansprüche. Die Dokumente und Daten in der Domäne „Biotech“ können nur durch die Benutzer der Gruppe „Biotech“ gesichtet werden, und diejenigen der Domäne "Biotech Mgmt" stehen ausschliesslich dem Leiter „HeadBio“ zur Verfügung.

Jede Domäne hat zwar ihre eigene Benutzer- und Kollektionsverwaltung; es wird jedoch vorausgesetzt, dass alle Benutzer und Gruppen auch in der Hauptdomäne existieren. Der Administrator einer Subdomäne kann einer Auswahl von diesen Benutzern und Gruppen beliebige Rechte für seine Subdomäne erteilen. Jeder Benutzer kann mehreren Domänen angehören, und er kann unterschiedliche Rechte in den verschiedenen Domänen haben.

Schutz der Dokumente

Der Zugriff auf ein Dokument wird nur dann gewährt, wenn die folgenden drei Bedingungen gleichzeitig erfüllt sind:

- Der Benutzer muss einer Gruppe angehören, die Zugriff auf die entsprechende Kollektion und ihre Datenquellen hat.
- Der Benutzer muss Zugriff auf die InfoCodex-Domäne haben, in der die Kollektion liegt.
- Der Benutzer muss Leserechte für das Dokument im zugrunde liegenden Betriebssystem haben, falls die File System Security aktiviert ist.

Falls die Dokumente einer Kollektion aus mehreren Quellen stammen, ist zusätzlich eine Beschränkung der Zugriffsrechte pro Quelle möglich.

3. Benutzeradministration und Zugriffsrechte

Die Benutzerverwaltung wird über die Menüpunkte System-Administration und „U1 – User-Daten“ bzw. „U2 – User-Gruppen“ aufgerufen.

3.1 Benutzer und Gruppen

Die Berechtigungen können für einen Benutzer in jeder InfoCodex-Domäne, in der er registriert ist, individuell gesetzt werden.

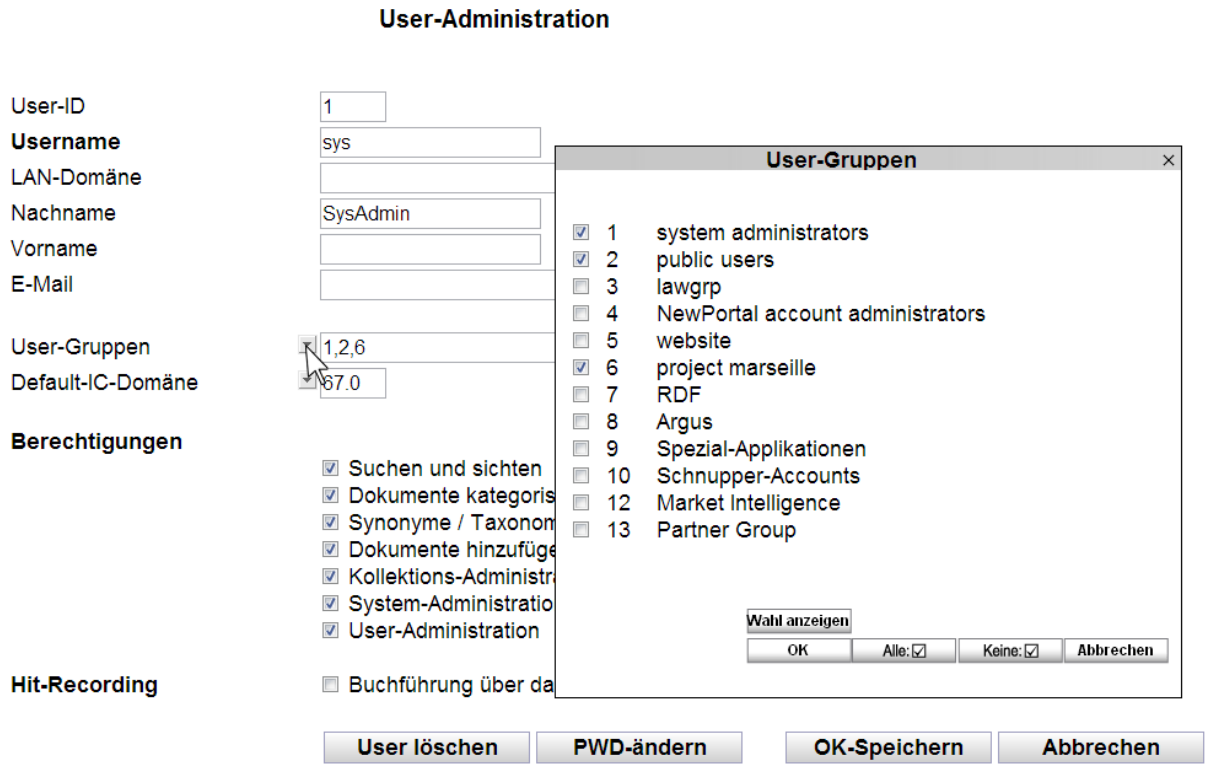


Abb. 3: Benutzerdaten und Gruppenzugehörigkeiten

3.2 Autonome InfoCodex-Domänen eröffnen

Die Eröffnung einer neuen InfoCodex-Domäne kann nur durch einen Systemmanager erfolgen, d.h. durch einen Benutzer, der in der Hauptdomäne über das Recht "User-Administration" verfügt.

Bei der Eröffnung wird der neuen Domäne eine Domänenadministrator („Besitzer“) zugewiesen, der anschliessend über die volle Souveränität über diese Domäne verfügt. Selbst der Systemmanager kann nicht auf diese IC-Domäne zugreifen, wenn er vom Domänenadministrator nicht ausdrücklich dazu ermächtigt wird. Der Systemmanager kann hingegen die gesamte Domäne wieder löschen.

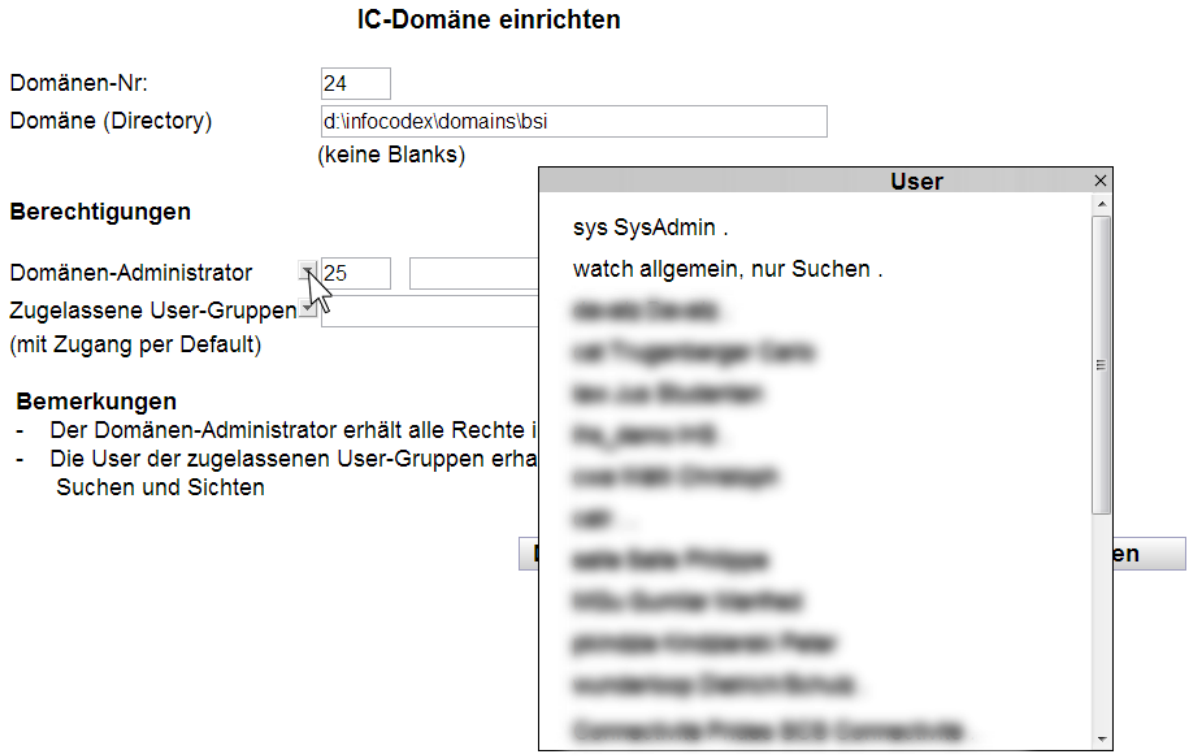


Abb. 4: Eine neue Domäne einrichten und den Administrator auswählen

3.3 Zentrale Benutzerverwaltung mit LDAP

Falls für die zentrale Benutzerverwaltung ein LDAP-Server eingesetzt wird (z.B. Active Directory oder Lotus Domino), können die aktuellen Benutzerkonten und die Gruppen periodisch übernommen und in InfoCodex automatisch nachgeführt werden.

Es ist empfehlenswert, auf dem LDAP-Server eine spezielle Benutzergruppe einzurichten (z.B. "InfoCodex"), welcher alle InfoCodex-Benutzer angehören. Der Systemadministrator muss dann lediglich diese Gruppe auswählen, um alle zugehörigen Benutzer sowie die Namen aller Gruppen, denen mindestens einer der Benutzer angehört, in InfoCodex zu importieren.

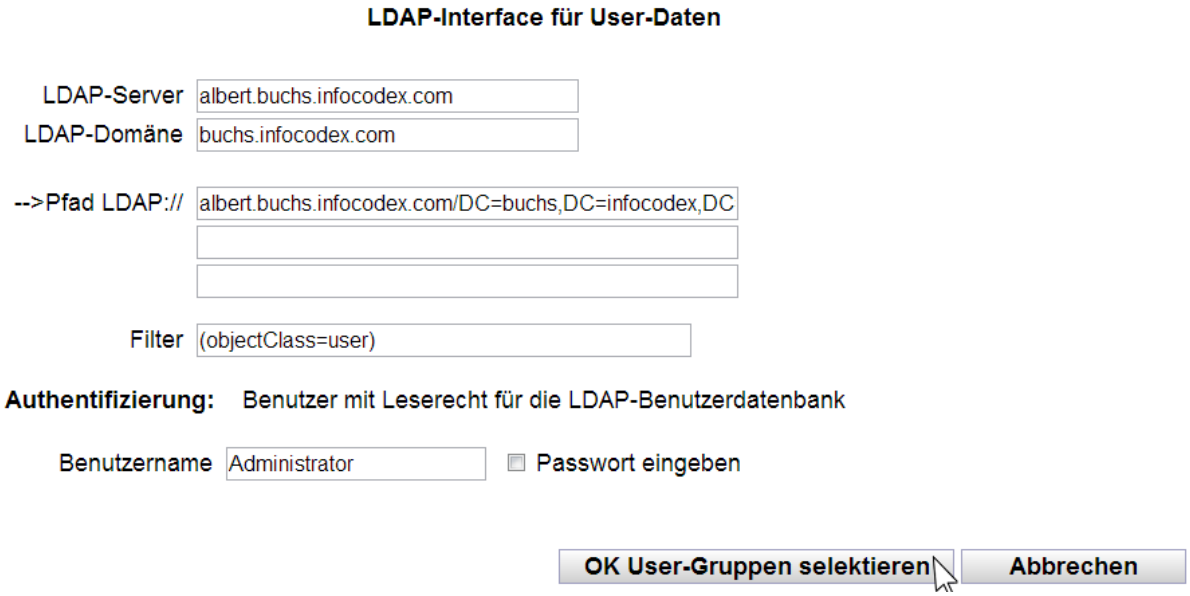


Abb. 5: Import von Benutzern und Gruppen über LDAP

Wirkungen eines Imports von Benutzerkonten durch LDAP:

- Neue Benutzer und deren Gruppen werden in die InfoCodex-Benutzerverwaltung aufgenommen.
- Benutzer, die im zentralen LDAP-Server nicht mehr vorkommen, werden in InfoCodex entfernt.
- Für bereits registrierte Benutzer werden die Gruppenzugehörigkeiten aktualisiert. Die in InfoCodex vergebenen speziellen Berechtigungen bleiben erhalten (ein Domänenadministrator behält beispielsweise seine Privilegien innerhalb seiner InfoCodex-Domäne).
- Neue Gruppen, denen mindestens einer der InfoCodex-Benutzer angehört, werden in die Datenbank von InfoCodex aufgenommen.

Um die im LDAP-Server vorgenommenen Änderungen wirksam zu machen, muss das Schnittstellenprogramm regelmässig aufgerufen werden. Alternativ kann das Programm ICLDAP2 durch den Job-Scheduler periodisch ausgeführt werden.

Hinweis: Unter Umständen ist es notwendig, dem InfoCodex-Account zusätzliche Berechtigungen zu erteilen, um via LDAP übernommene Benutzer korrekt zu authentifizieren. Dies ist dann der Fall, wenn beim Login die Fehlermeldung "Passwort ungültig" erscheint, auch wenn das Passwort richtig eingeben wurde. Es handelt sich dabei um die Privilegien "Einsetzen als Teil des Betriebssystems" und "Erstellen eines Tokenobjekts".

3.4 File System Security

Die Option "File System Security" (FS-Security) in InfoCodex bewirkt, dass die Zugriffsrechte des zugrunde liegenden Netzwerks konsequent beachtet werden. Dies ist eine zusätzliche Schutzmassnahme parallel zu den Sicherheitsmechanismen von InfoCodex.

Wenn diese Option aktiviert ist, kann der InfoCodex-Benutzer höchstens diejenigen Dokumente finden und sichten, auf die er unter dem zugrunde liegenden Netzwerk ohnehin zugreifen kann.

Da diese Überprüfung bei grossen Suchergebnissen das Netzwerk stark beanspruchen kann, lassen sich verschiedene Sicherheitsstufen einstellen, um einen guten Kompromiss zwischen Sicherheit und Performance zu finden.

Sicherheitsstufe	FS-Security	Anzeige von geschützten Dokumenten in der Trefferliste	Anzeige der totalen Trefferzahl in der Trefferliste
0	nicht aktiviert	ja *	ja
1	nicht aktiviert	nein	ja
2	nicht aktiviert	nein	nein
3	aktiviert	ja *	ja
4	aktiviert	nein	ja
5	aktiviert	nein	nein

** Die geschützten Dokumente werden in der Trefferliste mit * markiert; sie können jedoch nicht gesichtet werden.*

Beim Import von Dokumenten, d.h. beim Aufbau der Kollektionen, bewirkt die Option, dass nur diejenigen Dokumente gelesen und analysiert werden, für welche der ausführende Benutzer die benötigten Leseberechtigungen hat. Dies bedeutet, dass ein Benutzer keine geschützten Dokumente in Kollektionen aufnehmen kann.

Die Sicherheitsstufe wird durch den Systemadministrator systemweit gesetzt (vgl. Abschnitt 7.2). Für die einzelnen Kollektionen kann die Sicherheit partiell gelockert werden, wobei folgende Regeln gelten:

- Beim Import ("Dokumente hinzufügen") wird in jedem Falle die durch den Systemadministrator gesetzte Sicherheitsstufe beachtet.
- Beim Suchen und Sichten gilt die Kollektionsspezifische Sicherheitsstufe. (Im Normalfall sind die beiden Sicherheitsstufen identisch).

Besonders hohe Sicherheitsstufen wirken sich nachteilig auf die Performance aus. Bei grossen Kollektionen sollte deshalb die Wahl der Sicherheitsstufe sorgfältig geprüft werden.

4. Kollektionen verwalten

4.1 Kollektionen einrichten und löschen

Diese Funktionen beziehen sich auf das Einrichten und die Pflege der einzelnen Kollektionen. Sie sind in Abschnitt 4.4 von Teil 1 des Benutzerhandbuchs beschrieben.

Hinweise:

- Falls der Import von Dokumenten durch Batch-Jobs gesteuert ist (vgl. Abschnitt 4.5 von Teil 1), können die entsprechenden Importskripte mit dem Button "C2 Datenquellen" gesichtet und allenfalls gelöscht oder editiert werden.
- Mit dem Löschen einer Kollektion (Button "C3 Kollektion löschen") werden alle Daten in der Kollektionsdatenbank von InfoCodex gelöscht. Die Dokumente, die zu dieser Kollektion gehören, bleiben natürlich unangetastet.

4.2 Vorbereitete Textdateien hinzufügen

Die Dokumente werden normalerweise mit der Funktion "Dokumente hinzufügen" in eine Kollektion geladen. Nach der Wahl der Datenquellen werden die selektierten Dokumente durch die Spider Agents gesammelt und temporär in ein einfaches Textformat konvertiert. Diese temporären Textdateien bilden die Basis für die folgende Inhaltsanalyse. Gleichzeitig erstellen die Spider Agents eine Liste der importierten Dokumente, in der die Dokumentnamen und die extrahierten Metadaten festgehalten werden.

Alternativ kann die Bereitstellung der Textdateien auch ausserhalb von InfoCodex durch andere Mittel erfolgen. Die Textdateien werden zu diesem Zweck in einem beliebigen Verzeichnis abgelegt. Zusätzlich muss eine Datei toc.lst erstellt werden, die die Namen und die optionalen Metadaten der zu importierenden Textdateien enthält.

Aufbau von toc.lst:

```
f2.txt|10.11.01|20|pdf|D:\widas32\gp\GPCIV.PDF|A.Meier|Aggregation by civ
f8.txt|24.06.01|32|pdf|D:\widas32\gp\GPSEX.PDF|A.Meier|Aggregation by sex
FName |Datum   |GP|Fmt|Quelldatei/-URL        |Autor  |Titel
```

"FName" bezeichnet den Dateinamen der Textdatei. Diesem Namen kann ein relativer oder absoluter Pfad vorangestellt werden.

"GP" ist eine Schätzung für den Grafikanteil im Quelldokument in Prozent.

Das Zeichen "|" wird zur Trennung der verschiedenen Datenfelder verwendet.

Das erste Feld (Name der Textdatei) ist obligatorisch. Die übrigen Felder sind fakultativ. Das Datum bezieht sich auf die letzte Änderung des Dokuments. Für das Dateiformat sind folgende Codes zu verwenden:

`doc`	Worddokument	`notes`	Lotus Notes
`xls`	Exceltabelle	`rtf`	Rich Text Format
`ppt`	PowerPoint-Präsentation	`text`	ASCII-Text
`pdf`	PDF-Dokument	`image`	Bilddatei
`html`	HTML-Datei	`xml`	XML
`email`	E-Mail	`ps`	PostScript
`msg`	Outlook-Element		

4.3 Informationslandkarte reorganisieren

Wenn Dokumente zu einer bestehenden Kollektion hinzugefügt werden (inkrementelles Laden), dann werden diese Dokumente den Feldern der Informationslandkarte zugewiesen, zu denen sie inhaltlich am besten passen. Die Struktur der Karte wird dabei nicht verändert.

Wenn der Anteil der nachträglich hinzugefügten Dokumente 20% übersteigt oder die hinzugefügten Dokumente ein Thema betreffen, das in der ursprünglichen Kollektion nicht vorgekommen ist, dann gibt die Informationslandkarte den Sachverhalt nicht mehr optimal wieder. In diesem Fall empfiehlt es sich, eine Reorganisation vorzunehmen.

Mit der Funktion "C4 Kategorisierung reorganisieren" werden die Dokumente nicht neu importiert und analysiert, sondern nur neu kategorisiert und indexiert. Die Verarbeitungszeit ist daher nicht allzu gross.

Für die Neugenerierung einer Kollektion mit neuem Import der aktuellen Dokumente steht die Funktion "C5 Neugenerierung der Kollektion" zur Verfügung (vgl. Abschnitt 6.2).

4.4 Verarbeitungsstatus einer Kollektion

Diese Funktion zeigt den Verarbeitungsstatus einer Kollektion sichten (speziell während Import und Analyse von Dokumenten). Ausserdem ermöglicht sie ein Sichten des Verarbeitungsprotokolls.

In der Maske "C1 Kollektion einrichten" (vgl. Abschnitt 4.4 von Teil 1 des Benutzerhandbuchs) wird der Status ebenfalls angezeigt, allerdings nur in Code-Form. Die Codes bedeuten:

0	noch keine Dokumente importiert
1	Import im Gange (interaktiv gestartet)
2	Import im Gange (durch Batch gestartet)
3	Import abgeschlossen; Inhaltsanalyse im Gange
-1	OK, Dokumente geladen und analysiert
-98	Index-Update erforderlich (nach Update der linguistischen Datenbank)
-99	Neugenerierung erforderlich (nach Update der linguistischen Datenbank)

In dieser Maske kann der Status im Bedarfsfall auch auf 0 zurückgesetzt werden, um die Kollektion in den Neuzustand zu versetzen.

5. Kategorisierung beeinflussen

Für die Beeinflussung der Kategorisierung von Dokumenten (Einordnung in ein sachlogisch aufgebautes "Büchergestell") und der Vergabe von Deskriptoren (Verschlagwortung der Dokumente) stehen folgende Mittel zur Verfügung:

Keyword-Setting

Deklaration einer Liste von Wörtern und Ausdrücken, die bei der Inhaltsanalyse besonders stark zu gewichten sind.

Vorgelagerte Datenbank

Bereitstellung einer kundenspezifischen linguistischen Datenbank (mit Links zu der Taxonomie), die 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 Informationslandkarte beeinflusst werden. In vielen Fällen kann mit der interaktiven und einfachen Ad-hoc-Kategorisierung (vgl. Kapitel 5.3) viel erreicht werden. Auf der anderen Seite besteht auch die Möglichkeit, fixe Kategorien vorzugeben und den Automatismus des selbstorganisierenden Neuronalen Netzes vollständig auszuschalten.

5.1 Keyword-Setting

Prinzip

Einer Liste von Wörtern/Ausdrücken kann ein besonders hohes Gewicht beigemessen werden (Stufen 2-, 4- und 8-fach). Diese Ausdrücke werden zusammen mit den gesetzten Gewichten auf einer Keyword-Tabelle gespeichert. Wird beim Einrichten einer Kollektion eine solche Keyword-Tabelle zugeordnet (vgl. Abschnitt 4.4 von Teil 1 des Benutzerhandbuchs), dann werden diese Ausdrücke bei der Inhaltsanalyse besonders stark gewichtet.

Das effektive Gewicht setzt sich aus der Signifikanz eines Wortes gemäss der linguistischen Datenbank von InfoCodex, dem Keyword-Gewicht und der Entropie des Wortes in der Gesamtkollektion zusammen.

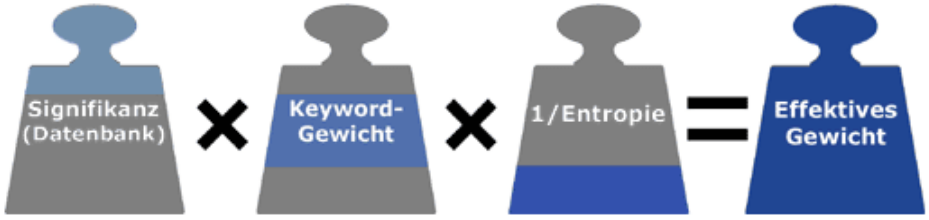


Abb. 6: Gewichtung von Ausdrücken in InfoCodex

Dabei bedeuten:

Signifikanz — Code in der linguistischen Datenbank für den Informationsgehalt eines Wortes/Ausdrucks im Bereich 0 bis 4:

0 unbedeutendes Füllwort, das in der Inhaltsanalyse ignoriert wird (z.B. "der", "in", "he")

...

4 sehr bedeutendes und klares Wort, z.B. "Weltgesundheitsorganisation"

Keyword-Gewicht	Spezielles Gewicht, das beim Keyword-Setting vergeben wird: 8 für sehr wichtige Wörter 4 für wichtige Wörter 2 für bedingt wichtige Wörter 1 für übrige Wörter (im Keyword-Setting nicht gesetzt)
Entropie	Mass für die Ungewissheit/Unbestimmtheit eines Wortes in der vorliegenden Dokumentenkollektion. Wenn das Wort "Reuters" in den meisten Dokumenten vorkommt, hat es eine grosse Entropie und trägt wenig zur Differenzierung der Dokumentinhalte bei. Deshalb erhält ein solches Wort ein kleines Gewicht.

Vorgehen zum Setzen von Keywords

Erster Schritt: Aufruf der Funktion "Keyword-Setting" und Wahl einer bestehenden Wortliste

Wahl:

- Relevanteste Wörter einer bestehenden Dokumentenkollektion, oder
- Wörter aus einem Bereich der InfoCodex-Datenbank, oder
- Wörter, die durch den Benutzer auf einer Textdatei bereitgestellt worden sind.

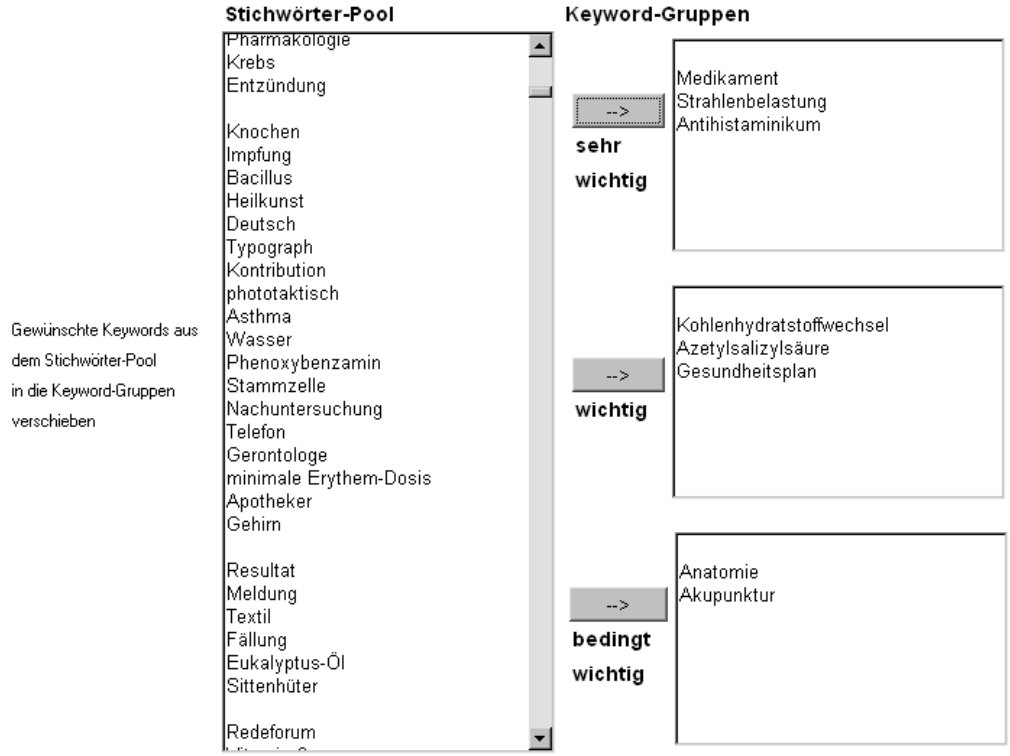


Abb. 7: Keyword-Gruppen

In der erscheinenden Maske können die als relevant betrachteten Wörter im Pool mit der Maus markiert und anschliessend durch Klicken eines der drei Buttons "sehr wichtig", "wichtig" bzw. "bedingt wichtig" einer Kategorie zugewiesen werden. Eine Direkteingabe von Wörtern ist ebenfalls möglich.

Beim Speichern werden die ausgewählten und eingegebenen Wörter zusammen mit den entsprechenden Gewichten in einer Keyword-Tabelle abgelegt. Der Tabellenname wird abgefragt; er muss mit ".kw" enden.

Die Keyword-Tabelle wird im Datenverzeichnis der aktuellen InfoCodex-Domäne gespeichert. Sie steht allen Kollektionen in dieser Domäne zur Verfügung.

Bestehende Keyword-Tabellen können auch gedruckt oder editiert werden. Die Gewichtung "0" bedeutet dabei "Entfernen aus der Tabelle". Für die Aufnahme zusätzlicher Wörter steht ein spezielles Eingabefeld zur Verfügung.

Die Liste kann nach Wörtern/Termen oder nach Gewicht sortiert werden. Die Buttons mit den Pfeilen dienen der Navigation innerhalb der Tabelle.

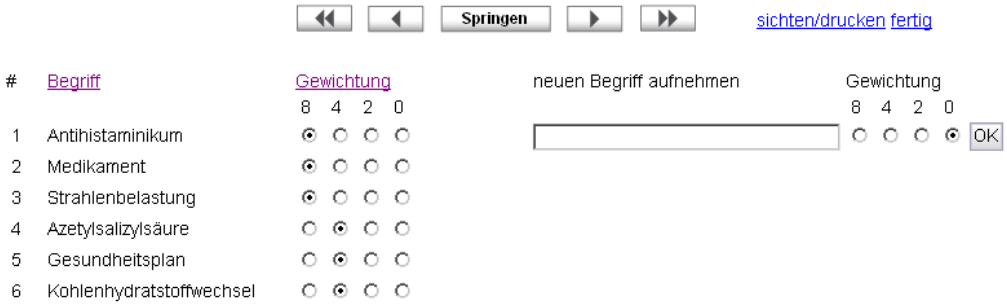


Abb. 8: Bestehende Keyword-Tabelle editieren

Zweiter Schritt: Zuordnung einer Keyword-Tabelle zu den Kollektionen

Mit der Funktion "Kollektion einrichten/löschen" kann einer Kollektion eine Keyword-Tabelle zugeordnet werden.

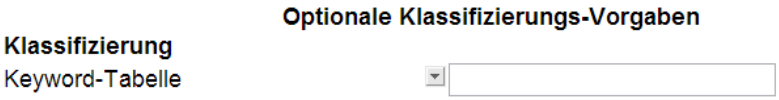


Abb. 9: Option „Keyword-Tabelle“ in der Maske „Kollektion einrichten“

Bei der Inhaltsanalyse von hinzugefügten Dokumenten werden die Wörter/Ausdrücke, die einem Eintrag in der Keyword-Tabelle entsprechen, besonders stark gewichtet. Dies beeinflusst die Kategorisierung sowie auch die Vergabe von Deskriptoren zu den einzelnen Dokumenten.

Falls eine Kollektion bereits besteht und die Zuordnung einer Keyword-Tabelle erst nachträglich erfolgt, muss nach der Zuordnung die Funktion "Reorganisation der Informations-Landkarte" aufgerufen werden.

5.2 Vorgelagerte linguistische Datenbank

Die bestehende linguistische Datenbank von InfoCodex umfasst mehr als drei Millionen Einträge und deckt praktisch alle Wissensgebiete ab. Sie stützt sich auf fundierte Werke wie das WordNet der Princeton University, EuroVoc, Agrovoc, Jurivoc, Taxonomien der Elektrizitätswirtschaft, der UBS, der Telekommunikation und vieler Fachverbände und UNO-Organisationen. Eine Anpassung an spezifische Problemstellungen ist daher nicht zwingend nötig.

Dennoch kann mittels einer fachspezifischen vorgelagerten Datenbank eine Qualitätssteigerung erzielt werden, die sich auch auf die Qualität der Deskriptoren auswirkt.

Beispiel: Die Abkürzung „SMD“ zeigt in InfoCodex auf "SMD-Bauelement" = "surface mounted device" → Komponente → Methode → usw.

In der Medienbranche steht SMD für "smd Schweizer Mediendatenbank".

In einer vorgelagerten linguistischen Datenbank kann der Begriff „SMD“ umdefiniert und zusammen mit "Schweizer Mediendatenbank" in die Synonymgruppe "smd Schweizer Mediendatenbank" gebracht werden. Diese Synonymgruppe zeigt auf den Taxonomieknoten „Name einer Organisation“ → Wirtschafts-/Geschäftszweig → Wirtschaft/Finanzen.

Die Begriffe in der vorgelagerten Datenbank haben immer Vorrang vor der linguistischen Datenbank von InfoCodex.

Bereitstellung einer vorgelagerten linguistischen Datenbank

Der Taxonomiebaum von InfoCodex besteht aus sieben Hierarchiestufen und umfasst ca. 4'100 Knoten. Jeder Knoten ist mit einem Hypernym (Oberbegriff) verknüpft, z.B.:

Windows 7 → Windows → Betriebssystem → Computer-Wissenschaft → Informatik → Information/Kommunikation.

Um zusätzliche Knoten im Taxonomiebaum einzuführen (z.B. für feinere Gliederungen in einem Fachgebiet), können die gewünschten Zusatzstrukturen in einer Exceltabelle bereit gestellt werden.

Diese Exceltabelle besteht aus 5 Textspalten zu maximal 30 Zeichen (Spalten C, D und E dürfen 34 Zeichen enthalten):

A	Mutterknoten (englisches Hypernym)	erforderlich
B	Tochterknoten (englisches Hypernym)	erforderlich
C	Tochterknoten (deutsches Hypernym)	fakultativ
D	Tochterknoten (französisches Hypernym)	fakultativ
E	Tochterknoten (italienisches Hypernym)	fakultativ
F	Tochterknoten (spanisches Hypernym)	fakultativ

Die englischen Hypernyme müssen eindeutig sein. Die in Spalte A (Mutterknoten) aufgeführten Hypernyme müssen entweder im InfoCodex-Taxonomiebaum bekannt sein oder in einer vorangehenden Zeile in der Spalte B deklariert sein.

A11

	A	B	C	D	E	F
1	**Newnodes.xls**					
2	**Declaration of New Nodes in the Taxonomy Tree**					
3						
4	**Known Hypernym**	**New Hypernym**	**New Hypernym**	**New Hypernym**	**New Hypernym**	
5	**(mother node)**	**(daughter node)**	**(German)**	**(French)**	**(Italian)**	
6						
7	reinsurance	reinsurance contract	Rückversicherungsvertrag	contrat de réassurance	contratto di riassicurazione	
8	reinsurance contract	reinsurance premium	Rückversicherungsprämie	tarif de réassurance	premio di riassicurazione	
9	reinsurance contract	reinsurance commission	Rückversicherungsprovision	commission de réassurance	commissione di riassicurazione	
10	reinsurance	reinsurance law	Rückversicherungsgesetz	droit des réassurances	legge sulle riassicurazioni	
11						
12						
13						

Abb. 10: Taxonomieknoten in Excel bearbeiten

Im obigen Beispiel ist "reinsurance" ein bestehender Knoten im Taxonomiebaum von InfoCodex:

ECONOMY/FINANCE	→	economic system	→	insurance	→	reinsurance
(Stufe 7)		(Stufe 6)		(Stufe 5)		(Stufe 4)

Die Deklarationen in der Exceltabelle bewirken eine Erweiterung des Taxonomiebaums um die vier unterstrichenen zusätzlichen Knoten:

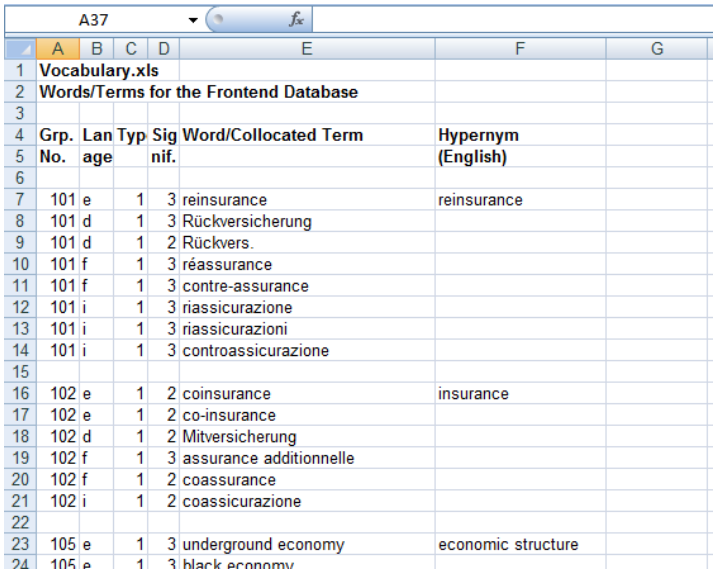
reinsurance (Stufe 4) → reinsurance contract (Stufe 3) → reinsurance premium (Stufe 2)

→ reinsurance commission (Stufe 2)

→ reinsurance law (Stufe 3)

Exceltabelle mit dem Vokabular

Der substanzielle Inhalt einer vorgelagerten Datenbank muss in einer Vokabulartabelle bereit gestellt werden. In dieser 6-spaltigen Excel-Tabelle sind die Wörter/Ausdrücke nach Synonymgruppen zusammengestellt und mit Qualifikatoren ergänzt.



	A	B	C	D	E	F	G
1	**Vocabulary.xls**						
2	**Words/Terms for the Frontend Database**						
3							
4	**Grp.**	**Lan**	**Typ**	**Sig**	**Word/Collocated Term**	**Hypernym**	
5	**No.**	**age**		**nif.**		**(English)**	
6							
7	101	e	1	3	reinsurance	reinsurance	
8	101	d	1	3	Rückversicherung		
9	101	d	1	2	Rückvers.		
10	101	f	1	3	réassurance		
11	101	f	1	3	contre-assurance		
12	101	i	1	3	riassicurazione		
13	101	i	1	3	riassicurazioni		
14	101	i	1	3	controassicurazione		
15							
16	102	e	1	2	coinsurance	insurance	
17	102	e	1	2	co-insurance		
18	102	d	1	2	Mitversicherung		
19	102	f	1	3	assurance additionnelle		
20	102	f	1	2	coassurance		
21	102	i	1	2	coassicurazione		
22							
23	105	e	1	3	underground economy	economic structure	
24	105	e	1	3	black economy		

Abb. 11: Vokabulartabelle mit Synonymgruppen (erste Spalte)

Die Vokabulartabelle muss wie folgt aufgebaut sein:

Spalte	Format	Inhalt
A	Numerisch	Synonymgruppe: Dies ist eine beliebige Zahl für die Identifikation von Synonymgruppen. Wörter/Ausdrücke mit gleicher Bedeutung (Synonyme) haben die gleiche Nummer.
B	Text	Sprache: d, e, f, i, s = deutsch, englisch, französisch, italienisch, spanisch; 0 = sprachunabhängig (z.B. Namen)
C	Numerisch	Worttyp: 1=Substantiv, 2=Verb, 3=Adjektiv, 4=andere
D	Numerisch	Signifikanz; siehe unten.
E	Text	Wort / Ausdruck
F	Text	Hypernym (englisch): Das Hypernym verknüpft eine Synonymgruppe mit einem Knoten im Taxonomiebaum. Es muss in englischer Sprache angegeben werden und einem Hypernym der InfoCodex-Taxonomie oder der zusätzlich eingeführten Knoten entsprechen. Bei fehlender Angabe wird der Synonymgruppe im Rahmen der Inhaltsanalyse keine Sinndeutung zugewiesen.

Faustregeln zur Festlegung der Signifikanz

Die Signifikanz ist eine Zahl zwischen 0 (bedeutungslos) und 4 (sehr wichtig und eindeutig), die bestimmt, wie bedeutend und aussagekräftig ein Begriff ist.

Typ	Default	1-2 Buchstaben oder unbedeutend	3-4 Buchstaben oder mehrdeutig	>16 Buchstaben oder bedeutend	Sehr wichtig und eindeutig
Substantiv	2	0	1	3	4
	Appartement	AG, sein	Bank	Kreditinstitut	Internationales Rotes Kreuz
Verb	1	0	0 oder 1	2	3
	atmen	werden	wird, warf	verstaatlichen	auf Vordermann bringen
Adjektiv oder andere	1	0	0 oder 1	2	3
	kommerziell	es, andererseits	gar, warm	mehrzellig	molekularbiologisch

Bestimmung des Gruppenführers

Das erstaufgeführte Wort in einer Synonymgruppe ist der Gruppenführer und sollte repräsentativ für die ganze Gruppe sein (pro Sprache). Es wird in InfoCodex als Deskriptor oder Neuronenlabel in Betracht gezogen und allenfalls angezeigt.

Wenn ein Wort eine um 5 erhöhte Signifikanz hat (z.B. 7 statt 2), dann wird dieses Wort zum Gruppenführer, auch wenn es nicht an erster Stelle steht.

Import der vorgelagerten Datenbank

Die bereit gestellten Excel-Tabellen müssen schliesslich als vorgelagerte Datenbank in das InfoCodex-System importiert und mit der linguistischen Datenbank abgeglichen werden. Dies geschieht im Adminbereich über den Menüpunkt „S3 – Vorgelagerte Datenbank“.

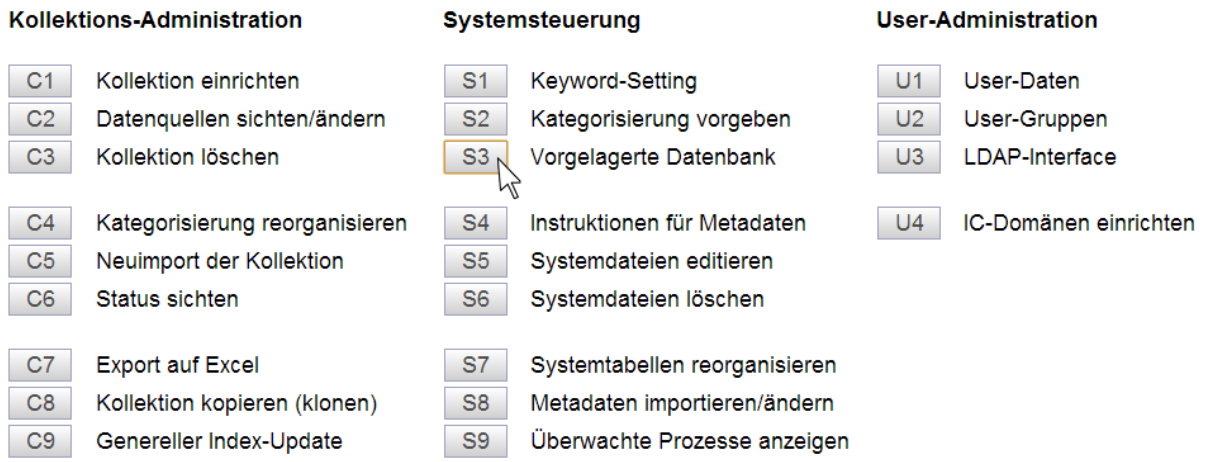


Abb. 12: Menu "S3 Vorgelagerte Datenbank" im Admin-Bereich

Beim Import wird der vorgelagerten Datenbank ein Name gegeben. Die vorgelagerte Datenbank wird im zentralen InfoCodex-Programmverzeichnis abgelegt. Sie steht allen Kollektionen in allen Domänen zur Verfügung.

Zuordnung einer vorgelagerten Datenbank zu einer Kollektion

Um einer Kollektion eine vorgelagerte Datenbank zuzuordnen, kann dies in den Kollektionseigenschaften explizit vorgemerkt werden:

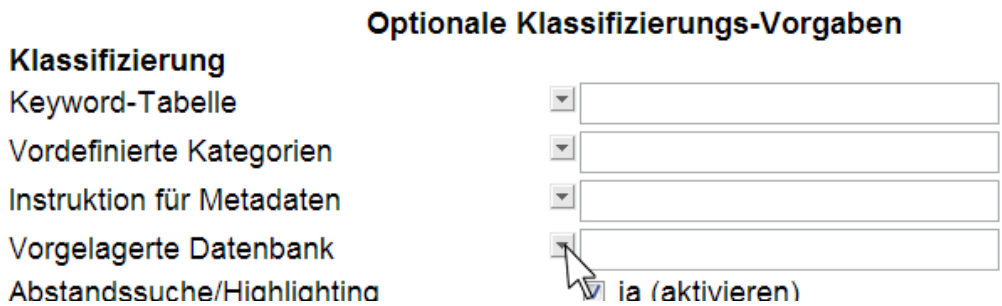


Abb. 13: Kollektionseigenschaften: Vorgelagerte Datenbank

Bei der nachträglichen Deklaration einer vorgelagerten Datenbank wird der Kollektionsstatus auf „Neugenerierung erforderlich“ gesetzt, und die Kollektion muss neu generiert werden (vgl. Abschnitt 6.2).

5.3 Vordefinierte Kategorien und Strukturen

InfoCodex kann die sachlogische Gliederung einer Dokumentenkollektion ohne menschliche Intervention vornehmen. Oft möchte der Benutzer aber die thematische Anordnung selber bestimmen. Dies ist zum Beispiel bei Response-Management-Applikationen der Fall, wo eine fixe Anordnung gefordert wird, um Antwortvorschläge aufgrund einer Entscheidungstabelle zu ermitteln.

InfoCodex bietet drei Möglichkeiten für die Erzwingung von vordefinierten Kategorien (Vorgaben für die Einteilung der Informationslandkarte):

- Ad-hoc-Kategorisierung — Einfache, durch Mausklicks gestaltbare Kategorisierung
- Analytische Kategorisierung — Wirksame, aber relativ aufwändige Vorgaben für die Gestaltung der Kategorisierung
- Fixe, trainierte Kategorisierung — Fixe Vorgabe und Trainieren der Kategorisierung mit Musterkollektionen

Im dritten Fall wird die Kategorisierung eintrainiert und bleibt als feste Vorgabe bei der Einordnung von neuen Kollektionen erhalten (d.h. das Neuronale Netz wird letztlich ausgeschaltet). Zur Förderung des Verständnisses für die drei verschiedenen Verfahren wird zunächst ein kurzer Überblick über die Kategorisierungsmethodik von InfoCodex gegeben.

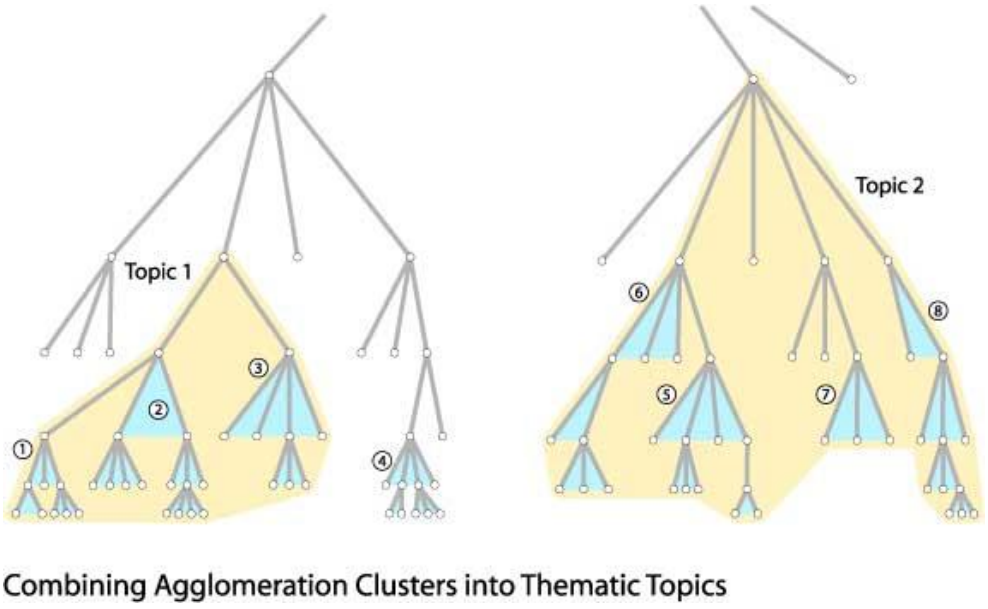


Abb. 14: [Erkennen von Themenschwerpunkten im Taxonomiebaum]

InfoCodex verfügt über eine universelle Taxonomie. Um die Dokumente einer Kollektion durch ein selbstorganisierendes Neuronales Netz optimal kategorisieren zu können, wird zunächst ein 100-dimensionaler Inhaltsraum konstruiert, der auf die gegebene Kollektion massgeschneidert ausgerichtet ist. Dabei werden diejenigen 98 Teilstrukturen im Taxonomiebaum ermittelt, die durch die Wörter in der Gesamtkollektion am meisten angesprochen werden (1 , 2 , 3 usw. in obiger Figur). Benachbarte Teilstrukturen werden anschliessend zu Haupthemen zusammengefasst (Topic 1, Topic 2 usw. in obiger Figur). Die 98 ermittelten charakteristischen Komponenten (Teilstrukturen des Taxonomiebaums) bilden zusammen mit einer Komponente für den Grafik- und Zahlenanteil eines Dokuments und einer Restkomponente die Koordinaten des 100-dimensionalen Inhaltsraums.

Die 98 charakteristischen Komponenten haben natürlich einen ganz entscheidenden Einfluss auf die Kategorisierung durch das Neuronale Netz. Die Beeinflussung der Kategorisierung setzt deshalb bei der Vorgabe oder teilweisen Vorgabe der charakteristischen Komponenten an.

Die zu einer Kollektion durch InfoCodex automatisch ermittelten Komponenten können mit dem Menüpunkt „S2 – Kategorisierung vorgeben“ → „Charakteristische Komponenten einer Kollektion sichten“ gesichtet werden.

Charakteristische Komponenten in der Dokumentenkollektion von "Pharma"

	Komponente	Stufe	Relevanz	Hauptthema
1.01	**Aktivität/Motiv**	6		
	Aktivität/Motiv	6	1234	
1.02	**Psychologischer Aspekt**	6		
	Psychologischer Aspekt	6	1431	
1.03	**Psychologischer Zustand**	6		
	Psychologischer Zustand	6	1207	
1.04	**Körperzustand**	6		
	Körperzustand	6	1051	Körperzustand
	Unordnung	5	1039	Körperzustand
	Gesundheit	5	500	Körperzustand
	Krankheit/Gebrechen	5	1446	Körperzustand
	Krankheit	4	1650	Körperzustand
	Infektionskrankheit	2	638	Körperzustand
	Hautkrankheit	3	440	Körperzustand
	Krebs	2	647	Körperzustand
	Infektion	5	764	Körperzustand
	Verwundung	5	464	Körperzustand

Abb. 15: Von InfoCodex automatisch ermittelte Themengebiete

Diese Liste zeigt insbesondere auf, welche Aspekte (Konzepte) mit welcher Relevanz in einer Kollektion hauptsächlich vertreten sind.

Variante 1: Ad-hoc-Kategorisierung

Diese einfache und anschauliche Kategorisierungsmöglichkeit wird über die Menüpunkte „S2 – Kategorisierung vorgeben“ → „Ad-hoc-Kategorisierung“ aufgerufen.

Nach der Wahl des Detaillierungsgrades für die Darstellung (empfohlen: 300 – 1‘000 Knoten) zeigt das System denjenigen Teil des Taxonomiebaums, der für die aktuelle Kollektion relevant ist. Die höchsten Hierarchiestufen sind links und die Detaillierung zu tieferen Stufen zieht sich nach rechts. Die Graustufen für einzelnen Bereiche zeigen die Relevanz für die gegebene Dokumenten-Kollektion auf: dunkelgraue Gebiete werden durch die Dokumente stark angesprochen, hellgraue Gebiete dagegen nur schwach. Die farbigen Flächen markieren diejenigen Teilbereiche des Taxonomiebaums, die als Hauptthemen für die Kollektion ermittelt worden sind.

WISSENSCHAFT/ KUNST/ LEHRE				
Wissenschaft/Forschung	Linguistik			
	Mathematik	Angewandte Mathematik		
	Naturwissenschaft	Biologie		
		Chemie		
		Erdwissenschaft	Geologie	
		Medizin	Pharmakologie	
			Medizinischer Begriff	
		Physik	Mechanik	
			Kernphysik	Radiologie
			Teilchenphysik	Elementarteilchen-Physik
	Psychologie			
	Forschung			
Naturwissenschaftler	Biologe			
	Chemiker			
Kunst	Musik			
Spezial-Fakultät				
Erforschung/Abklärung	Analyse			
	Ermittlung/Bestimmung	Bezeichnung	Diagnose	
	Untersuchung			
	Auskunft	Informationsbeschaffung		
		Nachrichten		

Abb. 16: Darstellung für die Ad-hoc-Kategorisierung

Die Kategorisierung kann beeinflusst werden, indem die farbigen Flächen modifiziert und den eigenen Vorstellungen angepasst werden. Folgende Operationen sind möglich:

1. Anklicken eines Knotens im grauen Bereich: der bei diesem Knoten beginnende und alle unteren Hierarchiestufen einschliessende Ast wird gefärbt und damit zu einem Hauptthema. Die Bezeichnung dieses Themas ist editierbar.
2. Anklicken eines Knotens in einem farbigen Bereich: Der bei diesem Knoten beginnende Ast wird herausgeschnitten und kann entweder inaktiviert werden (wird zur neutralen Grauzone), als eigenes neues Thema deklariert werden, oder mit einem anderen Thema verschmolzen werden.

Nachdem alle gewünschten Modifikationen durchgeführt worden sind, kann die neue Konfiguration als vordefinierte Kategorisierung gespeichert werden. Der Name dieser Kategorisierung wird abgefragt; er muss mit ".pm" enden ("predefined map"). Die gespeicherte PM-Tabelle steht allen Kollektionen in der aktuellen Domäne zur Verfügung.

Um die Ad-hoc-Kategorisierung tatsächlich umzusetzen, muss die Funktion „C4 – Kategorisierung reorganisieren“ ausgeführt werden.

Variante 2: Analytische Kategorisierung

Diese theoretisch gut fundierte Methode besteht aus der direkten Bestimmung der oben beschriebenen 98 charakteristischen Komponenten (Dimensionen des Inhaltsraumes). Sie wird über „S2 – Kategorisierung vorgeben“ → „Analytische Kategorisierung“ aufgerufen.

Schritt 1: Bestehende Komponenten-Selektion laden

Wahl der durch InfoCodex automatisch ermittelten Komponenten-Selektion als Vorlage.

Schritt 2: Überarbeiten der Komponenten-Selektion

Anhaken der als charakteristische Komponenten effektiv gewünschten Teilgebiete aus dem Taxonomiebaum (siehe abgebildetes Auswahlfenster). Mit dem Anhaken eines Knotens werden alle zugehörigen tieferen Hierarchiestufen eingeschlossen (104.05 umfasst alle bis

und mit 104.0511). Wenn hingegen ein Knoten auf einer tieferen Stufe zusätzlich angehakt wird, dann bildet dieses Teilgebiet eine eigene Komponente und wird aus dem höher liegenden Teilgebiet ausgeschlossen (in 104.08 ist im abgebildeten Beispiel der Ast 104.0801 ... ausgenommen; letzterer bildet eine eigene Komponente).

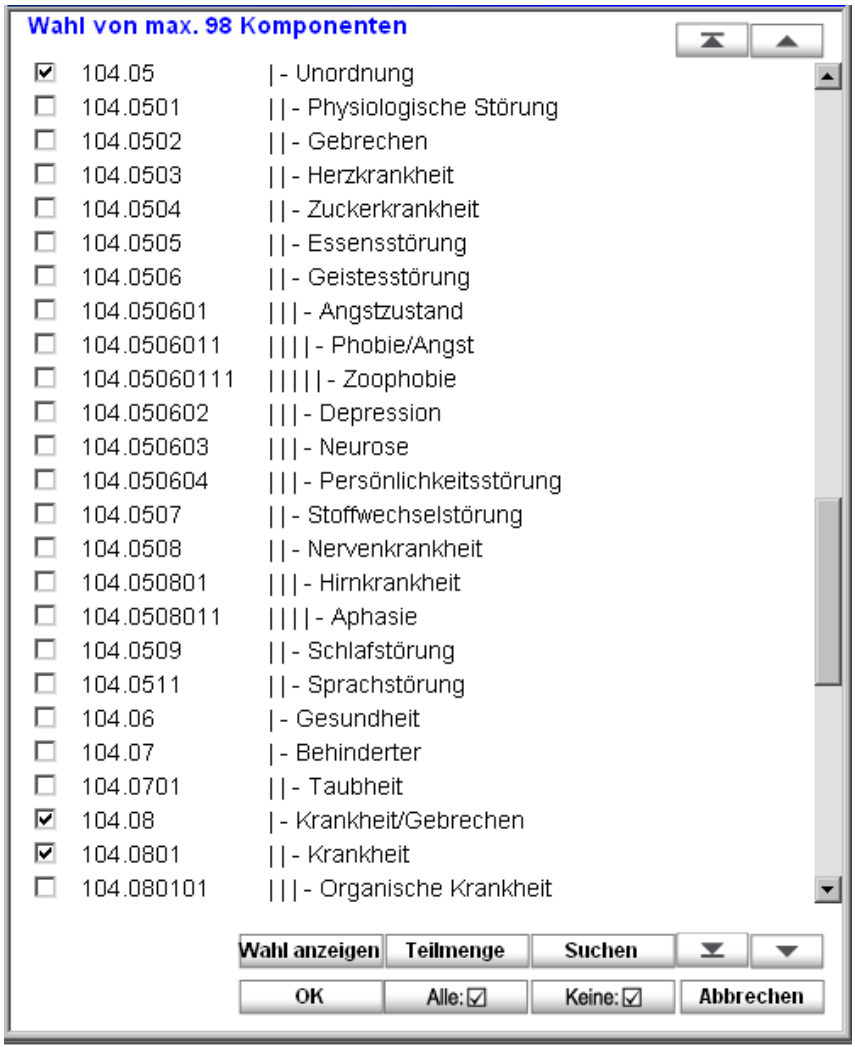


Abb. 17: Wahl der Komponenten

Wenn weniger als 98 Komponenten selektiert werden, erfolgt die Ergänzung auf 98 automatisch durch InfoCodex.

Schritt 3: Definition von Hauptthemen

Wahl von 2 bis 20 Hauptthemen, in welche die Informations-Landkarte zu unterteilen ist.

Schritt 4: Zuordnung zu den Hauptthemen

Zuordnung der 98 charakteristischen Komponenten zu den unter Schritt 3 gesetzten Hauptthemen.

Nach Schritt 4 wird die Kategorisierung auf einer PM-Tabelle gespeichert, und es muss anschliessend eine Reorganisation der Informationslandkarte erfolgen.

Variante 3: Fixe, trainierte Kategorisierung

Bei dieser Variante wird die Kategorisierung mit Hilfe von Musterkollektionen (Dokumente mit bekanntem Inhalt) trainiert. Den Dokumenten der Musterkollektion wird dabei mitgegeben, zu welchem Hauptthema in der Informationslandkarte sie gehören müssen.

Die Kategorisierung wird in diesem Fall in der Trainingsphase fix ermittelt. Beim späteren Importieren von Dokumenten bleibt die fixe Kategorisierung erhalten und der Automatismus des Neuronalen Netzes wird ausgeschaltet.

Vorgehen zur Bildung von fixen Kategorien:

Schritt 1: Bereitstellung einer Musterkollektion für das Training

Es sind 2 bis 20 Hauptthemen zu definieren, welche die gewünschte Zielstruktur der Karte beschreiben, z.B. "Monopole", "Steuern" und "Staatsbesitz". Die Dokumente einer einschlägigen Musterkollektion mit bekanntem Inhalt sind anschliessend derart auf separate Verzeichnisse aufzuteilen, dass die zum gleichen Hauptthema gehörigen Dokumente im gleichen Verzeichnis liegen (ein Verzeichnis pro Hauptthema mit mindestens je 20 Dokumenten).

Schritt 2: Erstellen eines Scripts für das Training

Die Instruktionen für die Bildung der vordefinierten Kategorien aus der bereitgestellten Musterkollektion müssen in ein Script eingetragen werden, das im Hauptverzeichnis der aktuellen InfoCodex-Domäne anzulegen ist, z.B. „C:\InfoCodex\data\instruct.txt“, wobei der Dateiname beliebig sein kann.

Das Script sollte folgendes Format haben:

```
Topic=monopoly
Dir=C:\demo2\sogei\monopoli
Topic=tax
Dir=C:\demo2\sogei\entrate
KW=vine, liquor, cigars

Topic=state property
Dir=C:\demo2\sogei\demanio
KW=castle, national park, railway
KW=*IDB: state property
KW=*UDB: state property
```

Dabei werden folgende Schlüsselwörter verwendet:

Topic — Die Namen der Hauptthemen müssen einem englischen Hypernym (Knoten im Taxonomiebaum) entsprechen. Als Hypernyme sind auch die Zusätze aus einer vorgelagerten Datenbank zugelassen.

Dir — Die Dokumente dieses Verzeichnisses (inkl. Unterverzeichnissen) sollen für das entsprechende Hauptthema repräsentativ sein. Diese Dokumente werden für das Trainieren der Kategorisierung benutzt.

KW — Optionale Schlüsselwörter:

„KW=Wein, Likör, Zigarren“ bedeutet: In der Trainingsphase sind die Schlüsselwörter "Wein, Likör, Zigarren" zu jedem Dokument in diesem Verzeichnis hinzuzufügen.

„KW=*IDB: state property“ bedeutet: In der Trainingsphase sind alle Wörter aus der InfoCodex-Datenbank mit Hypernym "state property" zu den Dokumenten hinzuzufügen. (Analog für eine vorgelagerte Datenbank mit KW=*UDB: ...)

Die optionalen Schüsselwörter können die Zielstrukturen präzisieren.

Schritt 3: Trainieren des Netzwerks

Eröffnen einer neuen Kollektion), Eingabe der Kollektionsbezeichnung und Wahl von „Erweiterte Vorgaben“.

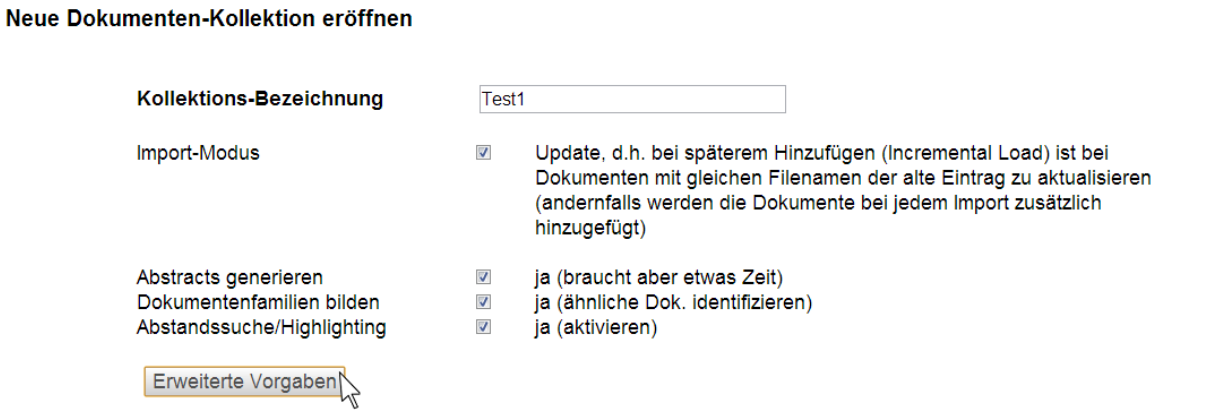


Abb. 18: Neue Kollektion eröffnen

Eingabe des Namens des erstellten Scripts mit einem führenden „*“ in das Feld „Vordefinierte Kategorien“, z.B. „*instruct.txt“. Es empfiehlt sich, die Anzahl Kolonnen für die zu generierende Karte vorzuschreiben, z.B. 20. Bei kleineren Musterkollektionen wird sonst die Dimension der Neuronenmatrix zu klein.

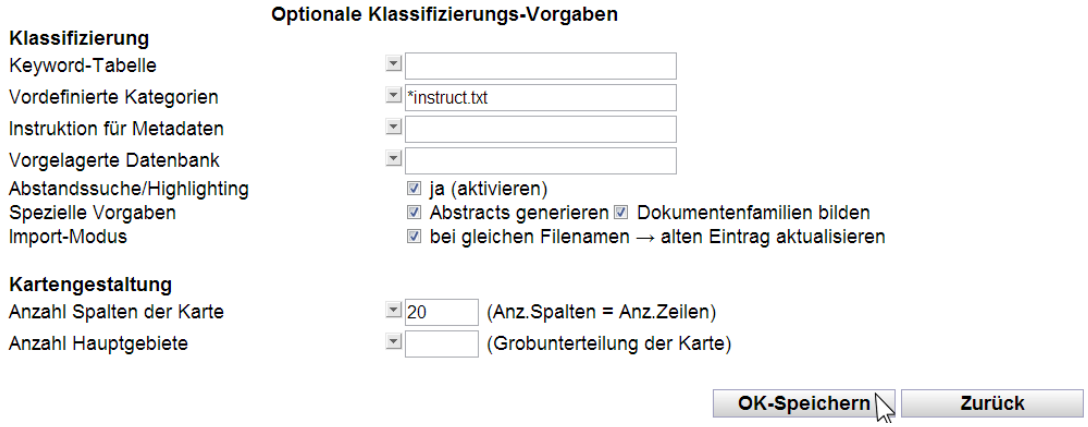


Abb. 19: Erweiterte Vorgaben – vordefinierte Kategorien und Grösse der Informationslandkarte

Das Trainieren beginnt unmittelbar nach dem Anklicken des Buttons „OK speichern“.

Schritt 4: Verbesserung der generierten Karte (optional)

Nach der Trainingsphase kann im Bedarfsfall noch eine Ad-hoc-Kategorisierung vorgenommen werden, um eine bessere Anpassung an die eigenen Vorstellungen zu erwirken. Anschliessend ist die Funktion „C4 – Kategorisierung reorganisieren“ auszuführen.

Schritt 5: Kollektionsstatus zurücksetzen und Dokumente hinzufügen

Die fixe Kategorisierung ist jetzt vollständig definiert, und es können beliebige Dokumente in diese fixe Struktur eingefügt werden. Das Neuronale Netzwerk bleibt unverändert.

Vor dem Import der Dokumente muss unter Menüpunkt „C1 – Kollektion einrichten“ noch der Status der trainierten Kollektion zurückgesetzt werden:

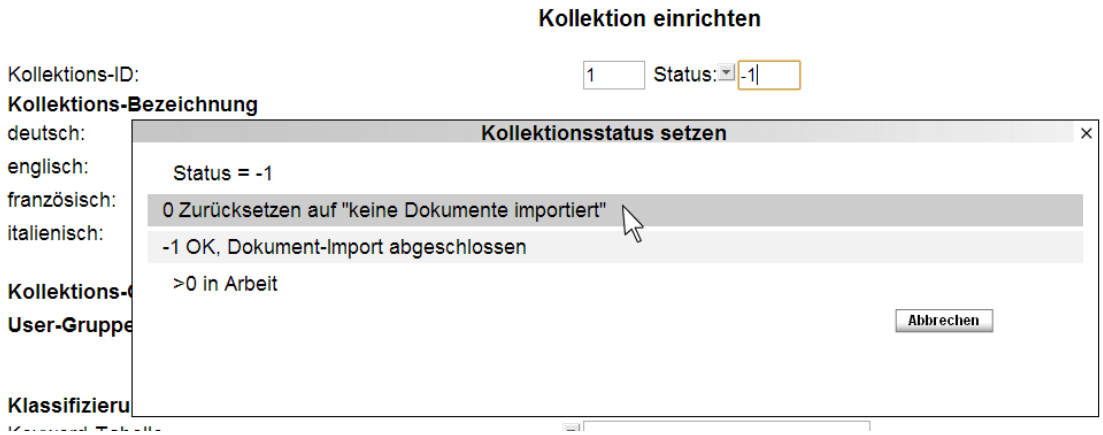


Abb. 20: Kollektionsstatus auf „keine Dokumente importiert“ zurücksetzen

Nach dem Zurücksetzen kann der Import von beliebigen Dokumenten in die trainierte Karte erfolgen.

6. Hilfsfunktionen

6.1 Import und Export von Exceltabellen

InfoCodex bietet diverse Möglichkeiten für einen Austausch von Serverdaten mit Exceltabellen auf dem Client. Dabei muss sichergestellt sein, dass das Plug-In, das von der InfoCodex-Startseite heruntergeladen werden kann, auf dem Client korrekt installiert ist.

Naheliegende Möglichkeiten sind die Ausgabe von Suchresultaten auf eine Exceltabelle und die Ausgabe von Kategorisierungsergebnissen und Verschlagwortungen (vgl. Abschnitt 6.1 von Teil 1 des Benutzerhandbuchs).

Weitere Möglichkeiten unter dem Menüpunkt „C7 – Export auf Excel“:

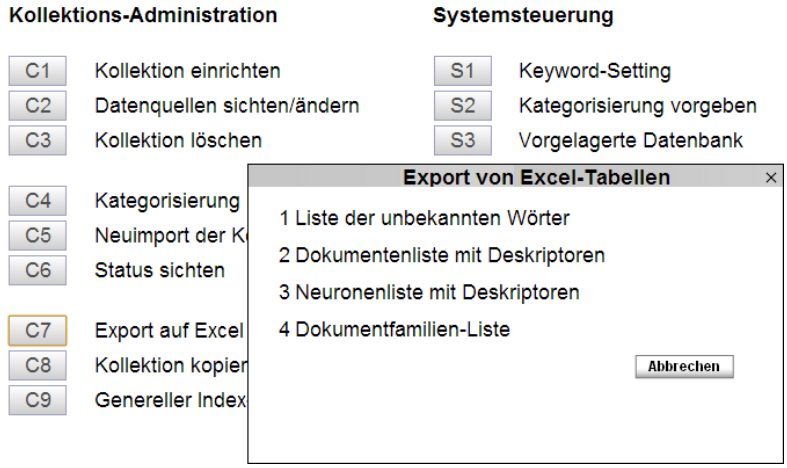


Abb. 21: Export auf Excel

Speziell erwähnt sei die Ausgabe von Wörtern, die in einer Kollektion vorkommen und in der InfoCodex-Datenbank nicht bekannt sind. Diese Liste kann einerseits dazu verwendet werden, Schreibfehler in den Dokumenten zu orten; anderseits kann sie als Basis für den Aufbau einer vorgelagerten linguistischen Datenbank benutzt werden.

6.2 Neugenerierung / Trefferstatistik

Diese Funktionen kommen insbesondere dann zum Tragen, wenn ein Update der linguistischen Datenbank erfolgt ist oder nachträglich eine vorgelagerte Datenbank eingeführt wird. Solche Änderungen haben zur Folge, dass die codierten Extrakte der Dokumente nicht mehr stimmen und neu konstruiert werden müssen.

Die verfügbaren Mittel ermöglichen eine Neugenerierung der Kollektionen, ohne dass die bisher erarbeiteten Import-Instruktionen, Keyword-Tabellen und vordefinierten Kategorisierungen verloren gehen.

Hit-Statistik

Dies ist eine laufend nachgeführte Statistik über das Anklicken der einzelnen Neuronen bzw. das Sichten von einzelnen Dokumenten pro Benutzer. Sie kann dazu dienen, im Bedarfsfall diejenigen Personen zu finden, die eine bestimmte Dokumentengruppe häufig konsultieren und daher als Experten auf dem entsprechenden Fachgebiet gelten könnten.

Bei einer Reorganisation bleibt die Hit-Statistik erhalten. Im Falle einer Neugenerierung wird der Benutzer gefragt, ob die Hit-Statistik neu initialisiert oder von der bestehenden Kollektion übernommen werden soll.

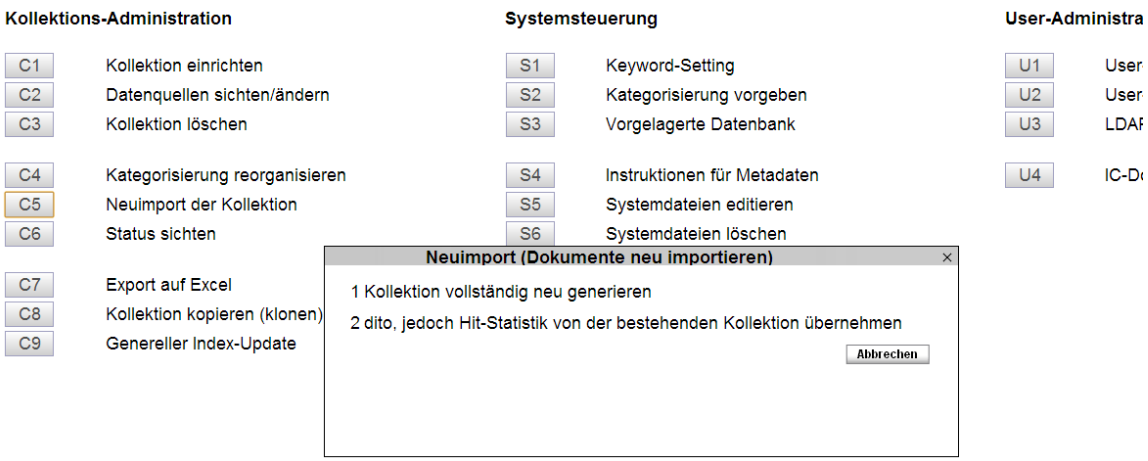


Abb. 22: Kollektion neu generieren; Hit-Statistik erhalten oder verwerfen

6.3 Kopieren, editieren, löschen

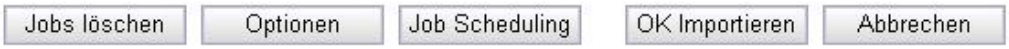
Menüpunkt „C8 – Kollektion kopieren“: Kopieren („klonen“) einer bestehenden Kollektion auf eine neue Kollektion

Menüpunkt „S4 – Systemdateien editieren“: Anpassen von Systemeinstellungen durch den Systemadministrator („options.ictxt“ usw.; vgl. Kapitel 7).

Menüpunkt „S6 – Systemdateien löschen“: Löschen von Keyword-Tabellen, vordefinierten Kategorien, vorgelagerten Datenbanken usw.

6.4 Job Scheduling

Wenn ein Import von Dokumenten periodisch durchgeführt werden soll, gibt man den Importauftrag zweckmässigerweise als Batch-Job für wiederkehrende Verarbeitungen auf (vgl. Kapitel 5.2 von Teil 1 des Benutzerhandbuchs):



Nach der Wahl „Job Scheduling“ erscheint eine Maske für die Spezifikation des Starts und der Periodizität der Verarbeitung.

Abb. 23: Optionen für periodischen Import

Unter "Gewünschte Aktion" wird angegeben, ob die Kollektion neu zu bilden ist („Initial Load“) oder ob die Dokumente zur bestehenden Kollektion hinzuzufügen sind („Incremental Load“).

Wenn Dokumente mit dem gleichen Dateinamen bei wiederholter Ausführung eines Importjobs nur einmal eingetragen werden sollen, ist beim Einrichten der Kollektion der Importmodus entsprechend zu setzen: „bei gleichen Namen → alten Eintrag aktualisieren“.

Wenn diese Option nicht aktiviert ist, werden die Dokumente bei jedem Import zusätzlich hinzugefügt (vgl. Abschnitt 4.3 von Teil 1 des Benutzerhandbuchs).

6.5 Monitoring und ständige Hintergrundprozesse

Die folgenden InfoCodex-Prozesse arbeiten ständig im Hintergrund:

Wimonsrv.exe

Wird auf Windows als Systemdienst installiert. Der Dienst startet und überwacht wimon.exe und verarbeitet Anmeldungsanforderungen von Benutzern, falls die Benutzerverwaltung von InfoCodex an eine Windows-Domäne angeschlossen ist.

Wimon.exe

InfoCodex-Scheduler, der unter Windows von Wimonsrv.exe bzw. unter Unix als Cronjob gestartet wird.

Dieses Programm wird pro Minute einmal aufgerufen und kontrolliert den Status der Kommunikations-Daemons wihjxd.exe für die WIDAS/HTML-basierte Administrationsoberfläche bzw. icapidaem.exe für das InfoCodex-API. Ausserdem prüft es alle 30 Minuten auf fällige Importjobs und startet diese gegebenenfalls.

ICRegen.exe

Wird von Wimon.exe periodisch zur Ausführung der fälligen Batch-Jobs aufgerufen.

Dieses Programm liest die aktiven periodischen Importjobs aus der Datei „importschedule.ictxt“, die im zentralen InfoCodex-Programmverzeichnis angelegt ist, und führt die nächsten Fälligkeiten in dieser Datei nach. Ausserdem fügt es die in der Datei „importqueue.ictxt“ durch die Benutzer neu aufgegebenen hinzu und entfernt gelöschte Jobs.

Die Periodizität des Aufrufs von ICRegen ist in der Datei „importnext.ictxt“ eingetragen; die Zahl in Zeile 1 ist das Zeitintervall in Minuten.

ICWatch.exe

Watchdog, der die aktiven Importprozesse überwacht und bei Problemen korrigierend eingreift, z.B. einen hängenden oder abgestürzten Prozess kontrolliert abbricht und an der abgebrochenen Stelle neu startet.

ICDelete.exe

Die unkoordinierte Löschung grosser Datenmengen (z.B. einer Kollektion mit einer Million Dokumenten) von Disk kann die Performance negativ beeinflussen und für den Benutzer unzumutbar lange Wartezeiten bedeuten. Darum werden diese Dateien nicht sofort gelöscht, sondern lediglich für die spätere Löschung markiert. ICDelete übernimmt die Löschung im Hintergrund, wenn der Server nicht anderweitig ausgelastet ist.

7. Systemeinstellungen

7.1 Proxyserver (proxy.ini)

InfoCodex liest beim Zugriff auf das Internet die Proxy-Konfiguration des auf dem Server installierten Standardbrowsers. Falls für InfoCodex eine abweichende Konfiguration verwendet werden soll, kann die Adresse des Proxyservers in der Datei „proxy.ini“ im InfoCodex-Programmverzeichnis abgelegt werden (typischerweise "C:\InfoCodex\pgm"). Die Datei hat folgendes Format:

1	`www-proxy.intern.local`	Hostname des Proxyservers
2	`8080`	Port
3	`Infocodex`	Benutzername (fakultativ)
4	`Secretpwd`	Passwort (fakultativ)

Bei fehlendem proxy.ini und Standardbrowser wird angenommen, dass kein Proxyserver eingesetzt wird und vom InfoCodex-Server direkt auf das Internet zugegriffen werden kann.

7.2 Verarbeitungsoptionen (options.ictxt)

Die Hintergrundprozesse von Infocodex können durch verschiedene Parameter kontrolliert werden. Es betrifft dies insbesondere die Spider Agents für den Import der Dokumente, das Text-Mining und die Kategorisierungs- und Indexierungsprozesse.

Die Optionen werden in einer zentral angelegten Datei "options.ictxt" im InfoCodex-Programmverzeichnis angegeben. Bei den Bezeichnungen wird nicht zwischen Gross- und Kleinschreibung unterschieden.

Format von options.ictxt:

```
SECURITY     = 4
RETARDANT    = 10
CLEAN        = 30000
SHOWMSG      = 1
OL_LocalPrf  =  infocodex-lokal
ExchangePrf  =  infocodex
```

Die möglichen Optionen sind nachfolgend zusammen mit dem Default-Wert aufgeführt:

SECURITY = 0

Spezifikation der Sicherheitsstufe (vgl. Abschnitt 3.4)

RETARDANT = 10

Die Hintergrundverarbeitungen, welche den Import und die Inhaltsanalyse der Dokumente besorgen, müssen auch bei grossen Dokumentenmengen zuverlässig zu Ende geführt werden. Gewisse Mehrprozessorsysteme haben unter Windows bei sehr hoher I/O-Last teilweise Probleme mit ihrem Cache-Management. Dies kann zu einem vorzeitigen Abbruch der Hintergrundverarbeitung führen.

Die Option RETARDANT bewirkt, dass die Hintergrundprozesse an kritischen Stellen eine kontrollierte Pause einlegen. Damit werden die erwähnten Probleme umgangen.

Empfohlene Werte:
- 0 für gut konditionierte Systeme
- 10 für Windows-Systeme mit sehr hoher Taktfrequenz oder mit 2 oder mehr Prozessorkernen
- 30 für Extremfälle

CLEAN = 30000

Das Text-Mining beansprucht grosse Mengen RAM. Als Vorsichtsmassnahme führt InfoCodex immer dann eine kontrollierte RAM-Bereinigung (garbage collection) durch, wenn die Anzahl der neu erkannten Wörter einen gewissen Grenzwert überschreitet.

Empfohlene Werte: 30000 für Systeme mit 256-512 MB RAM (=Default)

15000 für Systeme mit 128 MB RAM

50000 für Systeme mit mehr als 1 GB RAM

NUMBER = 20

Dokumente mit umfangreichen statistischen Tabellen (Messresultate, amtliche Statistiken) führen zu einer unverhältnismässig grossen Menge von verschiedenen Synonymen (jede Zahl entspricht einem anderen Synonym). Um den dadurch bedingten Overhead zu limitieren, kann die maximale Anzahl der in einem Dokument zu berücksichtigenden Zahlen durch die NUMBER-Option begrenzt werden.

Beispiele: 20 wenn nur die ersten 20 Nummern in einem Dokument von Interesse sind (Kundennummer, Referenznummer etc.)

1000 wenn bis zu 1000 Nummern pro Dokument interessieren

TRACE = 0

Wenn beim Import unerklärliche Probleme auftreten oder InfoCodex beim Import einer Website nicht alle erwarteten Links findet, kann die TRACE-Option zur Verfolgung des Imports aktiviert werden.

Beispiele: 0 kein besonderes Link-Tracing gewünscht

1 Minimale Debug-Informationen in ICAdd2.log festhalten

...

5 Äusserst detaillierte Debug-Informationen in ICAdd2.log festhalten

-2 Beim Web-Download sind alle festgestellten Links im Logfile „import.log“ festzuhalten

SHOWMSG = 0 (Microsoft Outlook)

Die Anzeige von E-Mails aus Outlook-Mailboxen, PST-Files oder MSG-Files, einschliesslich der Attachments, kann entweder im einfachen Textformat oder im Originalformat erfolgen. Dies wird durch die Option SHOWMSG geregelt. Für die zweite Variante muss Microsoft Outlook auf dem Client installiert sein.

Mögliche Werte: 0 Anzeige im einfachen Textformat

1 Anzeige im Originalformat (E-Mail und Attachments)

OL_LocalPrf (Microsoft Outlook)

Zu benutzendes Profil für das Outlook, das auf dem InfoCodex-Server installiert ist und zum Import von Mails aus PST-Files (Outlook-Archive) benutzt wird.

Beispiel: OL_LocalPrf = infocodex-lokal

OL_ExchangePrf (Microsoft Outlook)

Zu benutzendes Profil für das Outlook, das auf dem InfoCodex-Server installiert ist und zum Import von Mails vom Exchange-Server benutzt wird.

Beispiel: OL_ExchangePrf = infocodex

ExchangePwd (Microsoft Outlook)

Passwort für das Login auf dem Exchange-Server durch das Profil "OL_ExchangePrf" (sofern ein für "OL_ExchangePrf" effektiv ein Passwort benötigt wird).

OL_CleanPST = 0 (Microsoft Outlook)

Option für das forcierte Schliessen von offenen „Message Stores“ (PST-Files) vor dem Import eines PST-Files.

Mögliche Werte:
- 0 offene Message Stores sollen ignoriert werden
- 1 alle offenen Message Store” schliessen (ausser des Defaults)

ClientShare = \\infocodexserver\ICExchange$ (Microsoft Outlook)

Wenn ein Import von E-Mails aus der lokalen Outlook-Installation des Clients erfolgen soll, müssen die importierten E-Mails zwecks Inhaltsanalyse auf dem InfoCodex-Server zwischengespeichert werden. Diese Ablage erfolgt in einem temporären Unterverzeichnis des unter "ClientShare" spezifizierten Netzwerkpfads.

ClientDir = C:\InfoCodex\Exchange (Microsoft Outlook)

Dies ist der lokale Name des unter „ClientShare“ aufgeführten Verzeichnisses auf dem InfoCodex-Server.

TranslateToEN=0

Option für das automatische Übersetzen von Dokumenten, die nicht in einer der 5 InfoCodex-Standardsprachen (Deutsch, Englisch, Französisch, Italienisch, Spanisch) vorliegen.

Mögliche Werte:
- 0 nicht übersetzen
- 1 fremdsprachige Dokumente mittels Online-Übersetzer (Google Translate oder Systran) auf Englisch übersetzen. Primär unterstützt werden: Russisch, Chinesisch (traditionell oder vereinfacht), Arabisch, Indisch (Urdu, Hindi, Devangari), Japanisch, Portuguiesisch.

Achtung: Diese Option muss gesetzt werden, bevor eine Kollektion angelegt wird, um wirksam zu werden.

AccessLog=0

Die Protokollierung von Login-Versuchen (erfolgreiche und fehlgeschlagene) aktivieren.

Mögliche Werte:
- 0 nicht protokollieren
- 1 Alle Login-Versuche werden in der Datei icaccess.log protokolliert.

CompressFiles = 0

Die generierten Textdateien können im Bedarfsfall in komprimierter Form gespeichert werden. Die Komprimierung spart Speicherplatz und kann die Performance verbessern.

Diese Option wird nur bei Windows NTFS-Systemen unterstützt.

Mögliche Werte:
- 0 Keine Komprimierung
- 1 Generierte Textdateien (Pseudo-HTML-Files) werden komprimiert.

CutUArgs

Eine Leerzeichen-getrennte Liste von Parametern, die aus URLs entfernt werden sollen.

Manche Websites bauen Session-IDs oder andere flüchtige Informationen in ihre Hyperlinks ein (z.B. "http://www.example.com/page1.php?SESSIONID=12345678"). Bei jedem Download werden so neue URLs für dieselben Seiten generiert.

InfoCodex umgeht dieses Problem, indem in CutUArgs aufgeführte Parameter aus URLs entfernt werden, z.B.:

```
CutUArgs = PHPSESSID PHPSESSIONID ASPSESSIONID[A-Z0-9]+ RANDOM
```

MailServer, MailPort, MailFrom, AlertsFrom, MailAuth

Die Parameter MailServer und MailPort enthalten Namen oder IP-Adresse bzw. Port des SMTP-Servers, über den InfoCodex E-Mails verschicken soll. Der Parameter MailFrom wird als Absender verwendet (z.B. "MailFrom = InfoCodex-Server <infocodex@example.com>").

Wenn der Server auch für den Mailversand von Alerts verwendet wird, muss mit dem zusätzlichen Parameter AlertsFrom der Absender für die Alerts gesetzt werden (z.B. "AlertsFrom = InfoCodex-Alerts <alerts@example.com>").

Benötigt der MailServer eine Authentfizierung, so kann diese mit dem Parameter MailAuth konfiguriert werden (z.B. "MailAuth = myUser:myPassword").

TxtFiles = txt

Eine Leerzeichen-getrennte Liste von Dateierweiterungen, die zusätzlich zu *.txt als ASCII-Textfiles importiert werden sollen, z.B.:

```
"TxtFiles=log asc prn".
```

7.3 Schnittstelle zu Lotus Notes

Auf dem InfoCodex-Server müssen ein Notes Client und das InfoCodex-Modul „Notes Connect“ installiert werden.

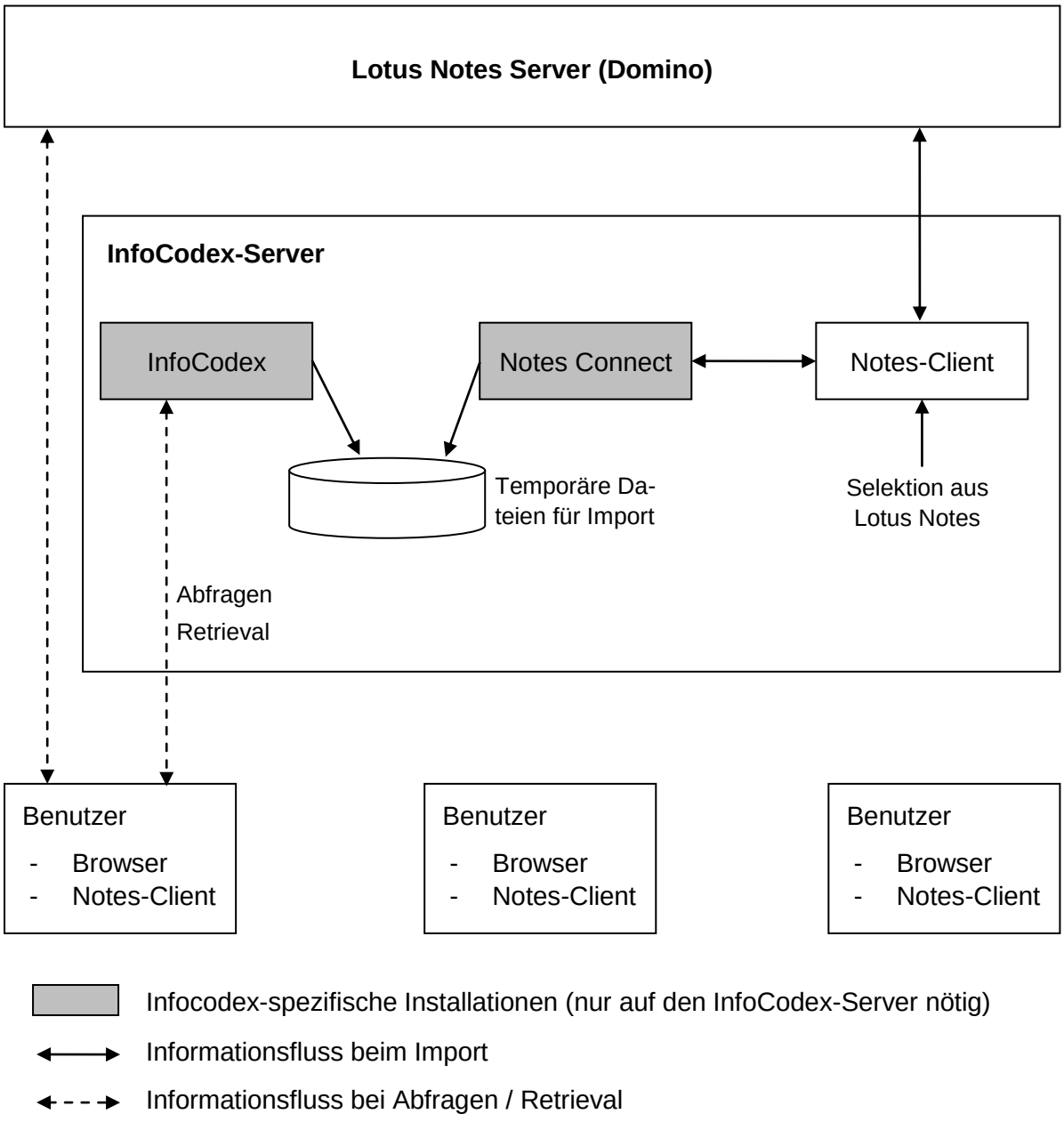


Abb. 24: Komponenten und Datenfluss beim Import von Lotus Notes

Import von Notes-Dokumenten

Die Selektion von Notes-Dokumenten und deren Attachments aus den verschiedenen Notes-Datenbanken erfolgt mit dem Notes-Client (dieser muss über die nötigen Berechtigungen verfügen).

Die selektierten Dokumente werden anschliessend in einem Hintergrundprozess auf temporäre Textdateien geschrieben, die nach der Analyse durch InfoCodex wieder gelöscht werden.

Die Notes-Dokumente stehen nun für Abfragen und Retrieval in InfoCodex zur Verfügung. Die Links zu den Notes-Datenbanken sind in InfoCodex abgelegt, so dass für die Anzeige

von einzelnen Notes-Dokumenten via Notes-Client auf die Originaldokumente zugegriffen werden kann.

Zugriffsberechtigungen

Beim Import der Notes-Dokumente können die Berechtigungen aus Lotus Notes mitgegeben werden. Die Koordination der Benutzerrechte mit Lotus Notes kann via LDAP erfolgen.

7.4 Schnittstelle zu Microsoft Outlook und Exchange Server

Für diese Schnittstellen muss Outlook auf dem InfoCodex-Server installiert werden, und es müssen die richtigen Einträge in "options.ictxt" gemacht werden (vgl. Abschnitt 7.2).

7.5 Dateiformate einschränken, externe Konvertierungsprogramme

Normalerweise bestimmt InfoCodex das Dateiformat anhand des Inhalts und ignoriert den Dateinamen. Damit können auch Dokumente korrekt erkannt und importiert werden, deren Dateierweiterung nicht mit dem Inhalt übereinstimmt (z.B. RTF in .doc, .xls/.xlsx/.csv vertauscht usw.). Allerdings kann dies bei grossen Mengen von nicht konvertierbaren Dateinamen den Import deutlich verlangsamen. In solchen Fällen kann der Dokumentenimport beschleunigt werden, indem man sich auf eine Auswahl von Dateierweiterungen beschränkt. In diesem Fall sind alle zu berücksichtigenden Erweiterungen in die Textdatei „extensions.ictxt“ entweder im zentralen InfoCodex-Programmverzeichnis (gilt generell) oder im Kollektionsverzeichnis (gilt nur für diese Kollektion) einzutragen.

Soll für eine bestimmte Datenquelle einer Kollektion ein eigenes „extensions.ictxt“ gelten, ist dies im Instruktions-File des Batch-Jobs (z.B. batch1.ins) von Hand zu setzen. Mit dem Schlüsselwort EXTENSIONS wird der Dateiname inkl. Pfad des quellenspezifischen „extensions.ictxt“ angegeben z.B. „EXTENSIONS=c:\xxx\myextension.ictxt“.

Beispiel einer Datei "extensions.ictxt":

```
doc,rtf,xls,xlt,ppt,pps,pdf,ps,psc,eps,msg,html,htm,shtml,xml,txt
tif,tiff,jpg,jpeg,jpe,bmp,png,gif,csv,tmp,asc,zip,gzip,gz
pst
```

Externe Konversionsprogramme

Zusätzlich zu den durch InfoCodex unterstützten Dateiformaten können beliebige Dateien indexiert werden, wenn ein entsprechendes Konvertierungsprogramm vorhanden ist. In der Textdatei „extensions.ictxt“ sind die Dateierweiterungen und der entsprechende Aufruf des Konversionsprogramm festzuhalten:

```
   ext1,ext2=myconvprg <in> <out> <meta>
```

Die Platzhalter <in> für Quelldatei (Originaldokument) und <out> für die zu erzeugende Textdatei sind zwingend notwendig. Falls das Konversionsprogramm zusätzlich Metadaten extrahieren kann, bezeichnet <meta> das entsprechende Metafile. Syntax und Schlüsselwörter des Metafiles sehen so aus:

```
AUTHOR: InfoCodex
SUBJECT: Gezielte Marktbeobachtung
TITLE: Erkennen Sie frühzeitig Markttrends und Fakten
MODIFIED: 12.04.2008
```

Wichtiger Hinweis:

Wenn von speziellen Konversionsprogrammen Gebrauch gemacht wird, dann müssen entweder alle zu importierenden Dateierweiterungen oder das Zeichen „*“ in „extensions.ictxt“ eingetragen werden.

Beispiel 1:

```
*
htm,html=c:\xxx\myhtmlconv <in> > <out>
```

Beispiel 2:

```
htm,html=c:\xxx\myhtmlconv <in> > <out>
```

Im Fall von Beispiel 2 werden ausschliesslich die Dateien mit den Erweiterungen ".htm" und ".html" importiert. Alle anderen Dateien werden übersprungen.

Import über iFilter

Ein IFilter ist ein externer Dateifilter, der Text aus einem bestimmten Dateiformat extrahieren kann. Er wird in der Regel vom Hersteller der Software mitgeliefert, die Dateien in diesem Format bearbeitet. InfoCodex kann verfügbare IFilter zum Import von Dateien verwenden.

Die Konvertierung mittels IFilter wird durch das Programm Filtdump.exe aufgerufen, das mit InfoCodex mitgeliefert wird.

Falls für spezielle Dateiformate ein IFilter vorliegt und diese Dateien auch indexiert werden sollen, ist dies in der Datei extensions.ictxt wie folgt festzuhalten:

```
dwg=filtdump  -b  <in>  >  <out>
```

7.6 Netzlaufwerke

Falls im lokalen Netzwerk bestimmte Laufwerksbuchstaben zu Netzwerkfreigaben zugeordnet sind, kann dies in der Textdatei „drives.ictxt“ im zentralen InfoCodex-Programmverzeichnis eingetragen werden. Der Benutzer sieht dann in InfoCodex seine gewohnten Kurznamen "R:", "S:" usw. anstelle der UNC-Pfade.

Beispiel einer Datei „drives.ictxt“:

```
S:->\\file002s\share
R:->\\file008s\share
Q:->\\file008s\pool
Z:->\\file008s\data\info_drive
N:->\\file008s\privat\<user>
P:->\\file002s\privat\<user>
```

Der Platzhalter „<user>“ bedeutet, dass es sich um benutzerspezifische Zuordnungen handelt. Er wird jeweils durch den Namen des aktuellen Benutzers ersetzt.

7.7 Erweiterte Daemon-Einstellungen (monitor.ictxt)

Die Datei monitor.ictxt enthält die Parameter für die ständigen InfoCodex-Hintergrundprozesse (vgl. Kapitel 6.5).

Einträge:

```
Cleanup 03:00
```

Das Programm Wimon führt täglich eine Aufräumaktion durch (Löschen von temporären Dateien, Neustart des Kommunikations-Daemons). Dieser Parameter legt den Zeitpunkt dafür fest.

```
DAEMDIR C:\Apache2\cgi-bin
APIDAEM  C:\InfoCodex\pgm\icapidaem.exe
```

Diese Einträge sind für den Betrieb der Daemons (HTML-Umgebung und API) nötig. Sie müssen und sollten normalerweise nicht verändert werden.

```
DAEMON C:\InfoCodex\pgm\icdelete.exe
DAEMON C:\InfoCodex\pgm\icwatch.exe
```

```
...
```

Das Schlüsselwort DAEMON bezeichnet Prozesse, die von Wimon überwacht und bei Bedarf neu gestartet werden sollen.

7.8 Konfiguration für den API-Daemon (webserver.ictxt)

Die Datei „webserver.ictxt“ enthält den Servernamen aus Sicht der Clients und das Stammverzeichnis des Webservers, z.B.:

```
demo.infocodex.com
C:\Apache2
```

7.9 Hardware-intensive Prozesse zeitlich beschränken

Um zu Spitzenzeiten wichtige Ressourcen (CPU, RAM, HD, Fileserver) für kleinere Prozesse und interaktive Benutzer freizugeben, können für grosse Kollektionen Sperrzeiten definiert werden. Während der angegebenen Zeitintervalle wird ein laufender Import unterbrochen und anschliessend fortgesetzt.

Zu diesem Zweck wird im Kollektionsverzeichnis eine Datei „pause.ictxt“ mit folgendem Format angelegt:

- 1 Zeile pro Zeitintervall
- Mit "#" oder ";" beginnende Zeilen und Leerzeilen werden ignoriert
- Format: Wochentage, Beginn, Ende der Sperrzeit, z.B.:

```
12345    08:00-17:30    # Mo-Fr zwischen 08:00 und 17:30 nicht arbeiten
12345--  08:00 - 17:30  # Gleichwertig, andere Schreibweise
1--4--7  22:00-06:00    # Zeitraum geht über zwei Tage (gültig)
```

Mehr Beispiele finden sich in der Datei pause.ictxt im InfoCodex-Programmverzeichnis. Diese Datei dient lediglich als Beispiel und hat keinen Einfluss auf das Programmverhalten.

8. Lizenzverwaltung

Bei der Installation von Infocodex wird die kundenspezifische Lizenznummer abgefragt. Um Missbrauch zu vermeiden, wird die Gültigkeit der Lizenz periodisch überprüft. Nach jeder erfolgreichen Prüfung wird die InfoCodex-Installation für einige Wochen oder Monate freigeschaltet. Nach Ablauf dieser Frist ist eine erneute Prüfung nötig.

8.1 Automatische Freischaltung

In der Regel ist eine Freischaltung für einige Wochen gültig. InfoCodex versucht einige Zeit vor Ablauf automatisch, die Freischaltung zu verlängern. Dazu muss eine HTTP-Verbindung zum Lizenzserver *lic.infocodex.com* aufgebaut werden können.

Schlägt die automatische Verlängerung wiederholt fehl, wird einige Tage vor Ablauf bei jedem Login eine Warnmeldung angezeigt.

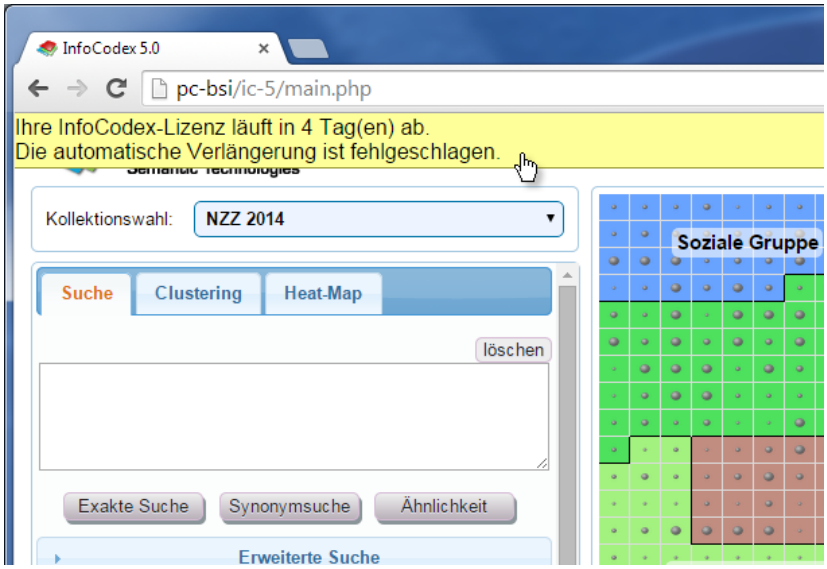


Abb. 25: Warnmeldung bei ablaufender Lizenz

8.2 Freischaltung/Lizenzerneuerung ohne Internetverbindung

Falls keine gültige Lizenz oder Freischaltung vorliegt und der Lizenzserver nicht erreichbar ist (z.B. weil der InfoCodex-Server keinen Internetzugang hat), wird beim Login eine entsprechende Fehlermeldung angezeigt.

Abb. 26: Keine gültige Lizenz vorhanden.

Um eine Lizenz ohne Internetverbindung freizuschalten, kann über die Menüpunkte "**Admin**" → "Systemadministration" → "**U5** – Lizenzerneuerung" notfallmässig eine Freischaltung in folgender Weise initialisiert werden:

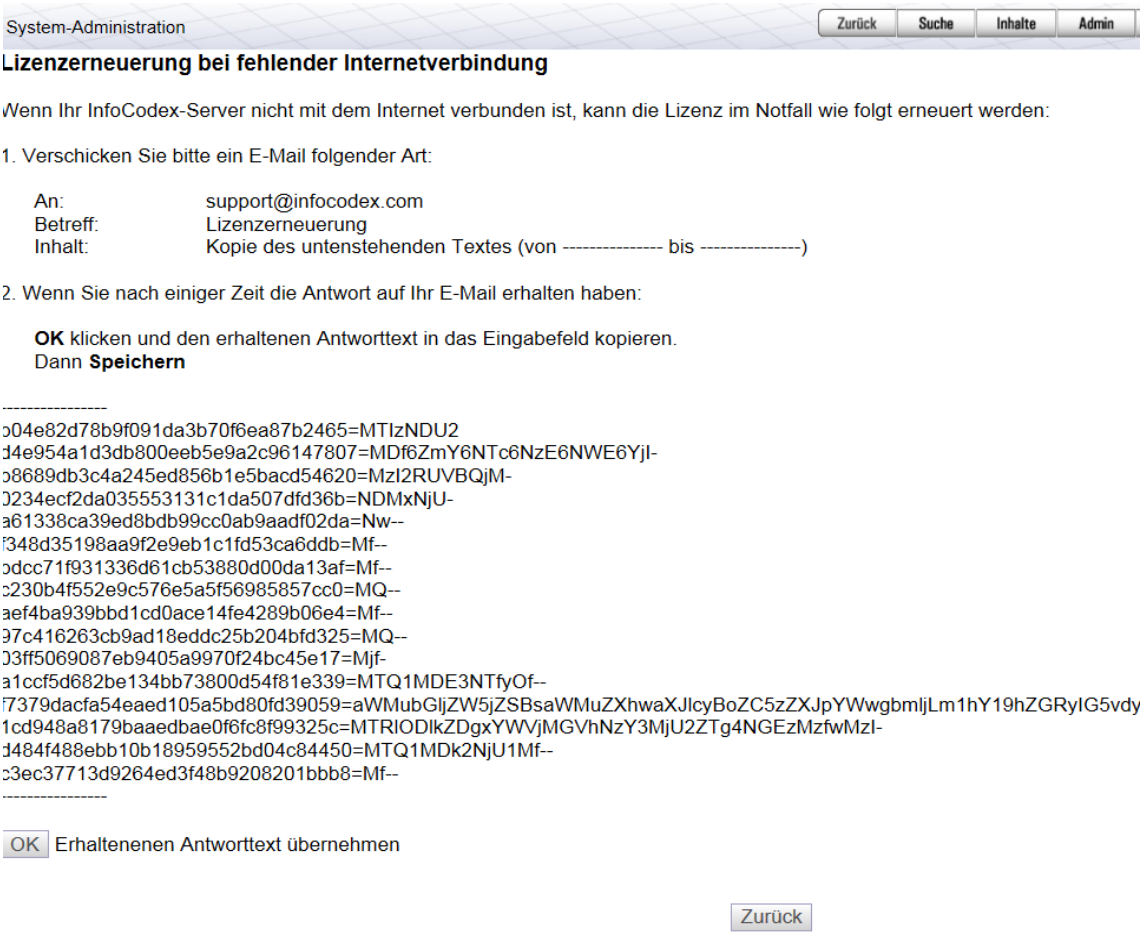


Abb. 27: Manuelle Freischaltung

Kopieren Sie den angezeigten Text von -------- bis --------- und senden Sie ihn an support@infocodex.com. Als Antwort erhalten Sie einen Freischaltcode, den Sie unter „OK – erhaltenen Antworttext übernehmen“ hinterlegen können. Damit wird der InfoCodex-Server freigeschaltet.