![InfoCodex Semantic Technologies logo]

User Manual

Part 2: System Management

List of Contents

1. Purpose of this Manual and Summary

Part 2 of the user manual is directed to system administrators and privileged users who have user and system administration rights for at least sub-domains. The following subjects will be outlined:

System Architecture and Data Protection Concept

•	*General*	Structure of the standalone and component versions; disk organization (software and linguistic database, frontend database, web interface); back-up recommendations
•	*Data Protection Concept*	Means to guarantee access protection to sensitive data; IC domain concept (autonomous regions); File system security

User Administration

•	*User data*	Creating/ removing users: Allocating privileges
•	*User groups*	Creating user groups to which users can be allocated
•	*LDAP interface*	Coordination of user data and user groups over a central user administration system, e.g. ADS from Windows (system administrator function)
•	*IC domains*	Establishing autonomous regions for particular user groups (system administrator function)

System Administration

•	*Collection administration*	Create/ remove document collections; Import procedures for associated data sources; Reorganization/ regeneration/ batch imports; Process status and log
•	*Regulation possibilities*	Ways to influence the categorisation (of the bookshelves) and the keyword setting
•	*Importing and Exporting Data*	Loading user-specific data from Excel spreadsheets; Outputing InfoCodex data to Excel spreadsheets

System Settings

•	*Processing control*	Setting process options; installation-specific environments (drive mapping etc.)
•	*Special interfaces*	Exchange Server; Lotus-Notes; Interfaces to special database systems

2. System Architecture and Data Protection Concept

2.1 System Architecture

Due to its open structure and portability, InfoCodex is relatively easily integrated into existing environments. The InfoCodex software is available in both standalone and in component form for integration into other systems.

Standalone-Version of InfoCodex

This is web-based and is installed on a host machine together with a standard web server (Apache or IIS). Client side, the only necessary software is a standard web browser.
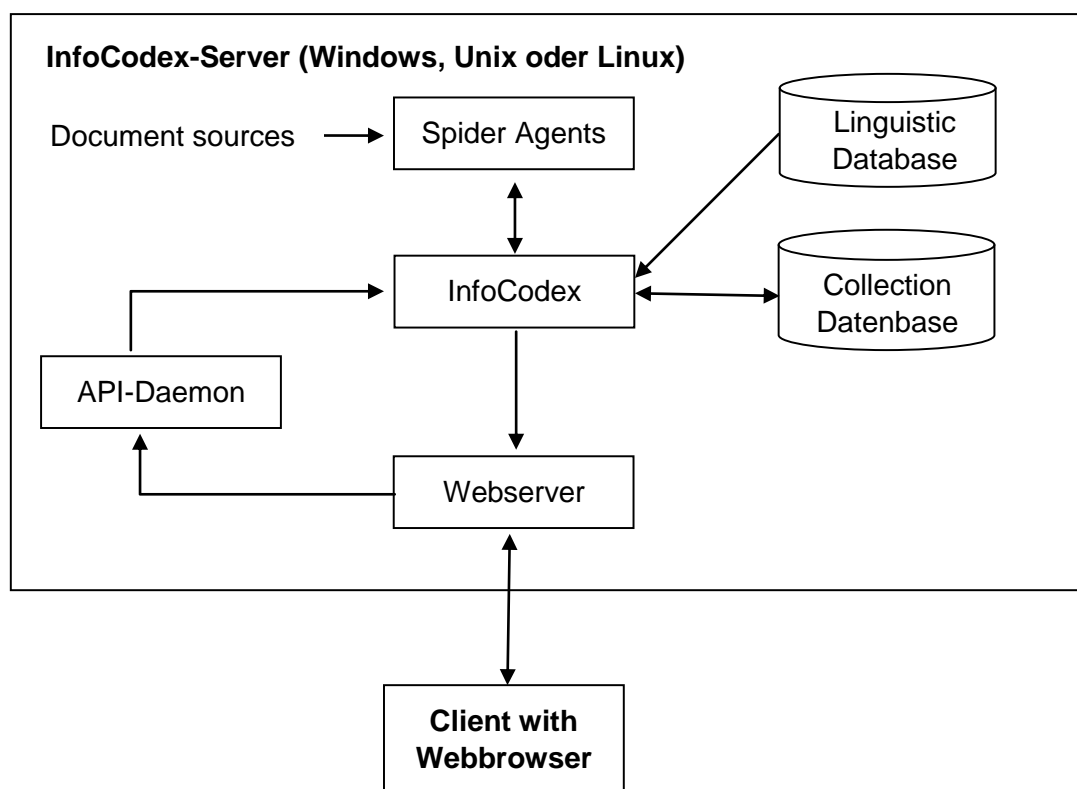
Fig. 1: Architecture of the standalone-version

Component Version of InfoCodex

InfoCodex can be embedded into other application systems in the form of its components. Due to its generic structure, user-specific solutions are easily implemented. At present there are interfaces for C, PHP, Perl and Java. The XML-based InfoCodex API is described in part 3 of the user manual.

2.2 Disk Organisation

The InfoCodex system is installed in the following three disk areas:

Program Area (stationary, requires no backup)

This area contains the software, the linguistic database, the global configuration files and several temporary files for process coordination.

The optional user-specific front end linguistic databases are also located in this area These are imported from Excel tables (= master data). They must be re-imported after each update of the InfoCodex linguistic database.

Data Area (user and collection data)

This contains the user databases, the local collection administration and, the individual collection databases.

It is recommended to include this data area in regular backup procedures.

Web Interface (*no backup needed*)

This area is used for the communication with the users' browser (HTML pages etc.).

Access rights fort he respective directories and files

Program area	wP		Data area	wP		Web interface	wP, r+
InfoCodex software	xP		User administration	wP		htdocs/icd	wP, r+
Linguistic database	rP		Collection administration	wP		htdocs/ice	wP, r+
Front end databases	wP					htdocs/icf	wP, r+
Process coordination	wP					htdocs/ici	wP, r+

r Read only		P InfoCodex Processes
w Read/Write		+ All ("world")
x Read and execute		

N.B.: When user-specific front end databases are used, the linguistic database must also have write access (wP).

2.3 Data Protection Concept

InfoCodex offers data protection at three different levels:

User Groups

Every InfoCodex user is a member of one or more user groups. Access rights to domains and collections are set per user and group. The actual access rights of a user are a product of his individual rights and of those groups to which he belongs.

File System Security

When this option is activated, a user can only view and find documents/files from data sources to which he would normally have access to dependent on the underlying operating system.

InfoCodex Domains

The InfoCodex data area wherein the collections are stored can be in one main domain and in an arbitrary number of sub-domains where each domain has its own user and collection administrator with individual access rights.

Access to a collection is dependent on the access rights to a specific domain in which the collection belongs. In this way, for example, a protected domain for members of a business management section could be set up so that not even the IT management would have access to it.
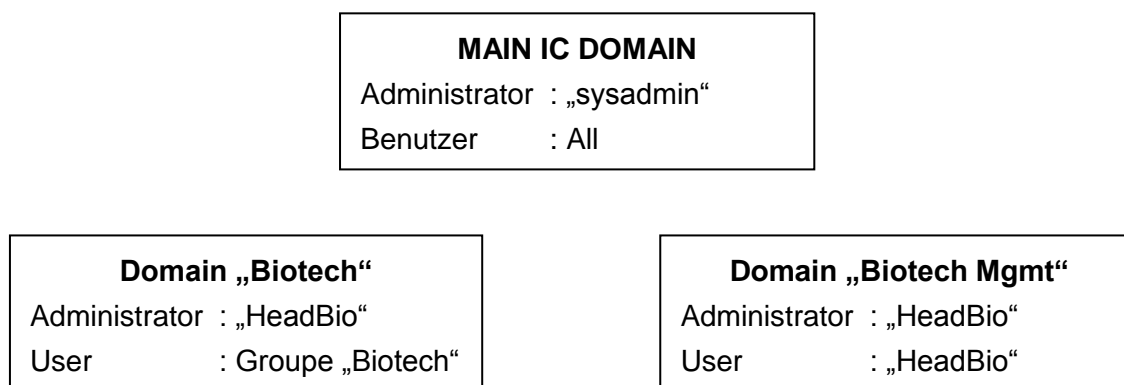
```
┌─────────────────────────────────────┐
│          MAIN IC DOMAIN             │
│  Administrator  : „sysadmin"        │
│  Benutzer       : All               │
└─────────────────────────────────────┘
```

```
┌──────────────────────────┐      ┌──────────────────────────────┐
│    Domain „Biotech"      │      │    Domain „Biotech Mgmt"      │
│ Administrator : „HeadBio"│      │ Administrator : „HeadBio"     │
│ User    : Groupe „Biotech"│     │ User        : „HeadBio"       │
└──────────────────────────┘      └──────────────────────────────┘
```

Fig. 2: InfoCodex domains with various user groups

All IC domains are created solely by the system administrator of the main domain. At creation time, a domain administrator is assigned who receives the sovereign rights for the corresponding IC domain. The domain administrator may allocate permissions to specific user groups or individual users. Even the system administrators of the main IC domain have no access to IC subdomains unless the permissions are explicitly given by the domain administrator. The system administrator has only the rights to delete the domain.

This concept fulfills all security requirements for the handling of confidential documents. The documents and data in the IC domain "Biotech" can be viewed only by the users belonging to the group "bio", whereas the documents in the IC domain "BioMgm" can only be viewed by the user "Headbio".

Each IC domain has its own user and collection administration. All users and all user groups must be registered in the main IC domain, i.e. the domain administrator can give permission only to users and user groups who are members of the user administration in the main IC domain. Each user can be a member of several IC domains, and he can have different access rights in the various IC domains.
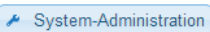
Document access

Access to a document is granted only under the following three conditions:

- The user belongs to a user group that has access to the corresponding collection and its data sources.

- The user is explicitly registered in the corresponding IC domain.

- The user has access to the document on the underlying file system level when File System Security is activated.

In the case where the documents in a collection come from several sources, it is possible to restrict access to particular data sources.

3. User Administration and Access Rights

This function is activated via the [⚲ System-Administration] button and then via the "U1 – User Data" resp. "U2 – User Groups" options.

3.1 User Data and User Groups

The user rights can be set individually for each domain in which the user is registered.

Fig. 3: User data and group membership

3.2 Creating IC Domains (Autonomous Regions)

Only the system administrators (users with "user administration" privileges over the main IC domain) can create a new IC domain.

A domain administrator is allocated on creation of a new IC domain, who then has complete souvereignty over that domain. Even the system administrator has no access to this domain unless expressly empowered by the domain administrator.

The system administrator does, however, have the power to delete any IC domains.

Fig. 4: Setting up a new domain and selecting athe domain administrator

3.3 LDAP Interface to a Central User Administration

The user administration of InfoCodex can be embedded into the general user administration of the underlying network by means of the LDAP standard.

Provided that an LDAP server is used for single point of administration (e.g. "Active Directory Service" or "Lotus Domino Server"), the current user data and their user groups can periodically be imported into InfoCodex from the central LDAP server.

It is recommended to introduce a special user group, e.g. "infocodex", assigned to all users that intend to use InfoCodex. The system administrator can then select all users belonging to the group "infocodex", and the interface program of InfoCodex will import exactly these users together with the names of the user groups occurring in conjunction with at least one user.

Fig. 5: Importing users and groups using LDAP

Consequences of a User Data Import by LDAP:

- New InfoCodex users and their user groups are inserted into the InfoCodex user administration.

- Users no longer present in the central LDAP server will be removed from InfoCodex.

- For the existing users, the user groups are updated. Higher access rights given to the users within the InfoCodex system will remain as they are (e.g. a domain administrator does not lose his privileges within his IC domains).

- New user groups, to which at least one of the InfoCodex users belongs, are imported.

To have the changes in the central LDAP server take effect for InfoCodex, the interface program has to be invoked (alternatively, the program ICLDAP2 can be periodically started by means of the job scheduler).

Note: In order to correctly authenticate LDAP users, InfoCodex may require additional privileges. This is the case if the login fails with the message "invalid password" even though the password is correct. If this happens, grant the privileges "Act as part of the operating system" and "Create a token object" to the user account under which InfoCodex runs.

3.4 File System Security

The activation of "File System Security" (FS security) in InfoCodex implies that the access rights set in the underlying network file system are consequently respected. This feature provides a supplementary protection measure parallel to the security mechanisms of InfoCodex.

The "File System Security" guarantees that the InfoCodex user can view at most those documents that he could access anyway on the underlying network file system.

Because of this access checking overhead in the case of large search results can strongly burden a network, there is the possibility of setting various security levels to achieve a good balance between performance and security.

Security Levels Offered by InfoCodex (see Section 7.2)

Security level	File System Security	Protected dokuments appear in the search result lists	Summary statistics shows the total number of hits
0	not activated	yes*	yes
1		no	yes
2		no	no
3	activated	yes*	yes
4		no	yes
5		no	no

*) The protected documents are masked with * in the search results list cannot be viewed.

With regard to the import of documents in the collection generation phase ("add documents"), the FS security option implies that only those documents are read and analysed, for which the user making the import has the required access rights. Therefore, a user cannot add protected documents inaccessible to his user account .

The FS security option is set system-wide by the system administrator. For the individual collections, the FS security can be deactivated. The following rules apply :

- At the import phase ("add documents"), the system-wide FS security set by the system administrator is respected.

- For the search and the document viewing, the collection dependent FS security is effective. (In the normal case, both security levels are identical).

The activation of the FS security has, of course, some impact on the performance. Its activation for collections with large amounts of documents should be carefully considered.

4. Managing Collections

4.1 Setting up / Deleting Collections

These functions involve the creation and maintenance of individual collections. They are outlined in Section 4.4 Part 1 in the user manual.

Note:

- In the case where documents are imported via batch job (see Section 4.3 Part 1), their scripts may be viewed or edited under the "C2 Data Sources" button.

- Deleting a collection (button "C3 Delete collection") removes all data from the InfoCodex collection database. The actual documents underlying the collection remain, of course, untouched.

4.2 Adding Prepared *.TXT File Collections

Usually, documents are loaded into InfoCodex with the function "Add documents" under the main menu "**Content**". After the selection of data sources, the InfoCodex spider agents gather the selected documents, convert them into a simple TXT file format and store those TXT files temporarily for the later use by the text mining programs. At the same time, the spider agents make a list "toc.lst" of the imported documents, in which the document names and the extracted metadata are registered.

The preparation of *.TXT files for import into InfoCodex can be performed independently by any other means. The documents must be deposited in an arbitrary directory and accompanied by a list file with the name "toc.lst", which contains the file names and optional metadata of the files to be imported.

a) Structure of 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  Date     GP Fmt Full name of the source Author  Document title
```

"fname" is the file name of the converted TXT file. It may be preceded by the path.

"GP" is an estimate for the graphics portion in the original document (in percent).

The "|" character is used as a field separator. The first field (name of the TXT file) is mandatory. The other fields are optional. Date refers to the last modification of the source file. For the format of the source file, use the following codes:

doc	Word document	notes	Lotus Notes
xls	Excel Spreadsheet	rtf	Rich Text Format
ppt	PowerPoint Presentation	text	ASCII-Text
pdf	PDF Document	image	Graphics file
html	HTML File	xml	XML
email	eMail	ps	PostScript
msg	Outlook Element		

4.3 Reorganising an Information Map

When new documents are added to an existing collection they are simply assigned to the existing fields of the information map where they fit best according to their content. These documents, however, have no influence on the structure of the collection map. It is necessary to perform a new reorganization of a collection in order to allow documents, added since the last classification / reorganization, to influence the structure of the information map. This is recommended when the number of newly added documents reaches somewhere between 10% and 20% of the original number of documents, or when documents from a new thematic area are added.

With the function "C4 Reorganise categorisation", the documents are not re-imported and analysed, they are only freshly categorised and indexed. The processing time is therefore not so significant.

To regenerate a collection with a fresh import of the current document set, use the function "C5 Reconstruct Collection" (see Section 6.2).

4.4 Processing Status of a Collection

This function displays the current processing stage of a collection's import and analysis phase. It also permits visual inspection of the processing log file.

The processing status can be viewed, in code form, under "C1 Setup collection". The codes have the following meaning :

0	no documents imported
1	Import in progress (interactively started)
2	Import in progress (started as batch)
3	Import complete; content analysis in progress
-1	OK, documents loaded and analysed
-98	Index update required (after an update to the linguistic database)
-99	Regeneration required (after changes to the linguistic database(s))

If necessary, the collection status can be set back to 0 in this input dialog point, in order to set the collection back to pre-import status, and a fresh import can be initiated.

5. Influencing the Categorisation

In order to influence the categorisation of documents ("shelving" in a thematic "book cabinet") and the allocation of descriptors, the following features are available :

Keyword setting   Declaration of a list of words or expressions that are to be heavily weighted during the content analysis phase

Frontend Database   Preparation of a customer-specific linguistic database (with links into the taxonomy tree), containing special technical expressions or company-internal abbreviations. A frontend database has priority over the standard Info-Codex database, such that during the content analysis, when a word is first found in the frontend database then <u>this</u> "meaning" and significance comes into effect.

Predefined Categories   This is a means to influence the grouping into the infor-mation map (i.e. the "book cabinet"). In many cases, the simple and interactive "<u>Ad-hoc Categorization</u>" can achieve very much. On the other hand, the possibility exists to set <u>fixed categories</u> and to deactivate the automatism of the neural network.

5.1 Keyword-Setting

Principle

A list of words or expressions (terms) can have an attached importance (weight) (levels 2, 4 and 8 times). These terms are stored in a keyword table together with their respective weights. Such a table can be assigned to a collection when it is set up (see Section 4.4 Part 1 in the user manual). These words / expressions then have a strong influence during the content analysis phase .

The effective weight is then determined from a combination of a word's significance (as re-gistered in the InfoCodex linguistic database), its keyword weight and its entropy over the whole document collection .
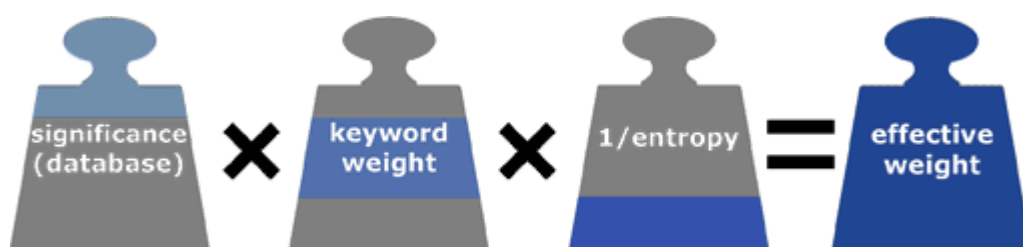
Fig. 6: Method of weighting of terms in InfoCodex

Explanation:

Significance	A code in the linguistic database for the information content of a word/ expression from 0 to 4 :

    0   meaningless "glue word" which is ignored in the content

        analysis phase (e.g. "the", "in", "he")

    ...

    4   very meaningful and unambiguous word, e.g.

        "World Health Organization"

Keyword weight   Special weight assigned via keyword setting :

    8 for very important words

    4 for important words

    2 for slightly important words

    1 others (not specified in keyword setting)

Entropy   A measure for the uncertainty / indeterminance of a word in the context of the current collection. When the word "Reuters" occurs in most of the documents, it has a greater entropy and contributes very little towards the differentiation of the documents' content.

    Such a word, therefore, receives a lesser weight.

Procedure for Setting Keywords

First step: Select the function "Keyword Setting" and choose an existing word list

Choices:

- Most relevant words of an existing collection, or

- Words from the realm of the InfoCodex database, or

- Words the user has provided in a text file.

Fig. 7: Keyword Groups

The relevant words in the pool are marked with the mouse and moved to the right by clicking the button indicating the desired relevance group (i.e. "very important", "important" or "not so important". Words can also just entered directly.

The saved keywords and weights are stored in a keyword table in the root directory of the current IC domain. The user is prompted for the table name which must have the extension ".kw".

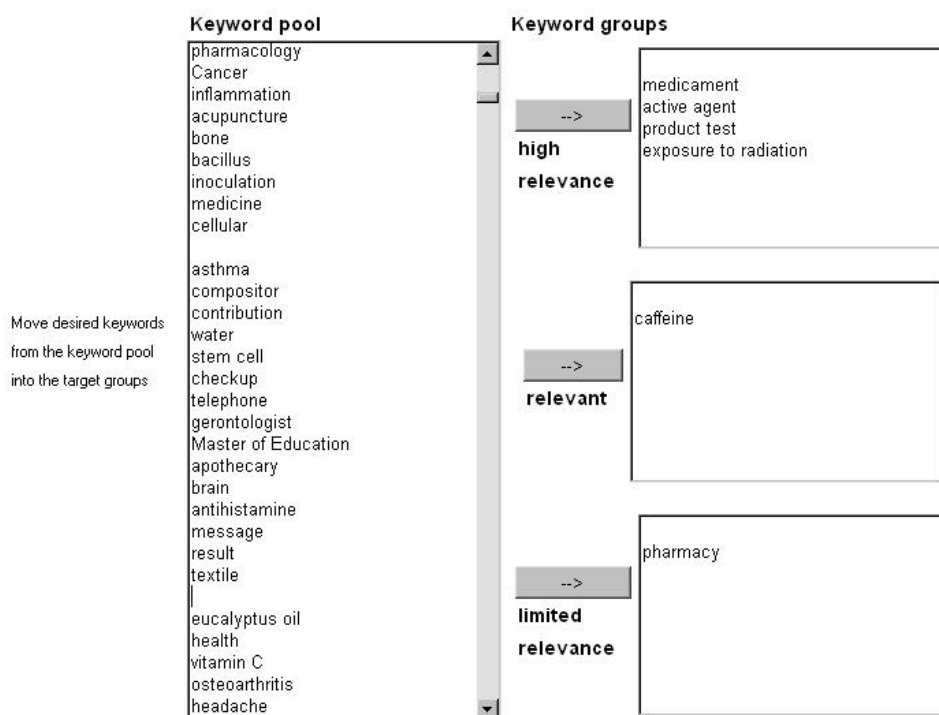The keyword tables are available to all collections in the current IC domain .

It is also possible to edit and print existing keyword tables. The assignment of words into keyword groups is achieved by checking the appropriate radio buttons.

Clicking the weight "0" means "removal from the keyword table .

The list can be ordered by term or by weight. Navigation through the table is performed with the buttons top of screen.

Fig. 8: Editing an existing keyword table

Second Step: Assigning a keyword table to a collection

A keyword table can be assigned to a collection via the function "set up / delete collection"

Fig. 9: Option: Allocate keyword table to a collection

During content analysis, the terms in the keyword table are weighted especially strongly. This influences both the categorisation and the allocation of descriptors of the individual documents.

If a keyword table is assigned after a collection has been created, then the function "reorganize information map" must be performed.

5.2 Front-End Linguistic Database

The current InfoCodex linguistic database consists of nearly 4 million terms and expressions and covers practically all areas of knowledge. It is based on established works such as WordNet from Princeton University, EuroVoc, Agrovoc, Jurivoc, Taxonomies from the electricity supply industry, UBS, telecommunication industry and many specialized associations and UNO organisations. An adaption for specialised environments is hardly necessary.

However, an increase in quality can be achieved by implementing a front end database which can have a significant impact on the quality of the document descriptors.

Example: InfoCodex interprets "SMD" as "surface mounted device" → component → method → etc

In the media branch, "SMD" stands for "Swiss Media Database ". By means of a front end database SMD can be redefined (placed in the same synonym group) as "Swiss Media Database" including the taxonomy link:

name of organization → economic branch → economy / finance

The terms and expressions in the front end database always have priority over those from the standard InfoCodex linguistic database.

Preparing a Front End Database

The InfoCodex taxonomy tree consists of seven hierarchic levels and encompasses about 4'100 nodes. Each node is identified by a hypernym (= superordinate term for the corresponding branch and all underlying branches in the taxonomy). For example:

Windows 7 → Windows → operating system → computer science → information → IN-FORMATION/ COMMUNICATION

If the user wishes to introduce additional nodes into the taxonomy tree (e.g. for finer structuring in a specialised area), the supplementary structures are prepared in an Excel spreadsheet on the client-side.

The Excel spreadsheet consists of 5 text columns of maximum 30 characters (columns C, D und E may contain 34 characters):

A	Mother nodes (English hypernym)	Obligatory field
B	Daughter nodes (English hypernym)	Obligatory field
C	Daughter nodes (German hypernym)	optional
D	Daughter nodes (French hypernym)	optional
E	Daughter nodes (Italian hypernym)	optional
F	Daughter nodes (Spanish hypernym)	Optional

The English hypernyms must be unique. The hypernyms in column A (mother nodes) must either be already present in the InfoCodex taxonomy tree or already declared in a previous row in column B.

	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						

Fig. 10: Defining taxonomy nodes in an Excel spreadsheet

In the above example "reinsurance" is a node present in the InfoCodex taxonomy tree

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

The declarations in the Excel spreadsheet introduce the following nodes as additional sub-structures:

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

                                              →   reinsurance commission
                                                  (level 2)

                 →   reinsurance law
                     (level 3)

Excel Spreadsheet with Vocabulary

The substance of the front end database must first be prepared in a vocabulary table. The terms in this 6-column table are allocated to synonyms groups and complemented with word qualifiers. It must have the following structure:

	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		

Fig. 11: Vocabulary table with synonym groups (see column 1)

The vocabulary table must be structured as follows:

Col	Format	Description		
A	N5	group number		
B	T1	language	e	= English
			d	= german
			f	= french
			i	= italian
			s	= spanish
			0	= *l*anguage independent (name, code)

C	N1	type (word qualifier)	1 = noun 2 = verb 3 = adjective 4 = others (adverb, pronoun)
D	N1	significance [b]	0 = meaningless ("glue word") 1 = not meaningful, but may be used as a search term in the synonym search 3 = important and unique 4 = very important and unique
E	T30	word / collocated term	e.g. European Court of Justice
F	T30	English hypernym [c]	

[a] Group number

This is an arbitrary but unique number for the synonym groups. Different terms with the same meaning (synonyms) must have the same group numbers. The term appearing first for a specific language is the group leader of the corresponding synonym group (unless the group leader is explicitly marked by a significance ≥ 5). The group leader is the most representative word of a synonym group; it might be displayed as a neuron or document label in InfoCodex.

[b] Significance

Range 0 to 4 according to the rules specified below. In order to mark the group leader of a synonym group, you may add the amount of 5 to the significance (range 5 to 9). If significance is not given, InfoCodex will generate a default significance.

[c] Hypernym

To link a synonym group to the taxonomy tree, a hypernym in the English language should be specified matching one of the nodes of the InfoCodex taxonomy tree or one of the user-supplied front end nodes. The hypernym entry is optional. The valid InfoCodex hypernyms may be output in Excel spreadsheet form.

Rules of thumb for significance

The significance is simply a number between 0 (meaningless) and 4 (very important and unique) that stipulates how meaningful and important an expression is.

Type	Default	1-2 letters or meaningless	3-4 letters or ambigous	>16 letters or meaninful	Very signifcant and unique
Noun	2 apartment	0 UN, check	1 bank	3 banking institution	4 nuclear power station
Verb	1 breathe	0 shall	0 or 1 come, sell	2 dehydrogen-ate	3 internationalise
Adjective or other	1 commercial	0 it, othetwise	0 or 1 well, lazy	2 multicellular	3 positively charged

Determining a Synonym Group Leader

The first term / expression of a synonym group we call the "group leader" which should be the most reprentative of the term involved. InfoCodex will represent this term as its descriptor or neuron label.

If a term has a significance raised by 5 (e.g. 7 instead of 2) then it will be treated as the group leader, even though it may not be on top of its group list.

Importing a Front End Database

The prepared Excel spreadsheet must be imported into the InfoCodex system and aligned with the standard linguistic database. See point "S3 – Front End Database" in the System Administration menu:
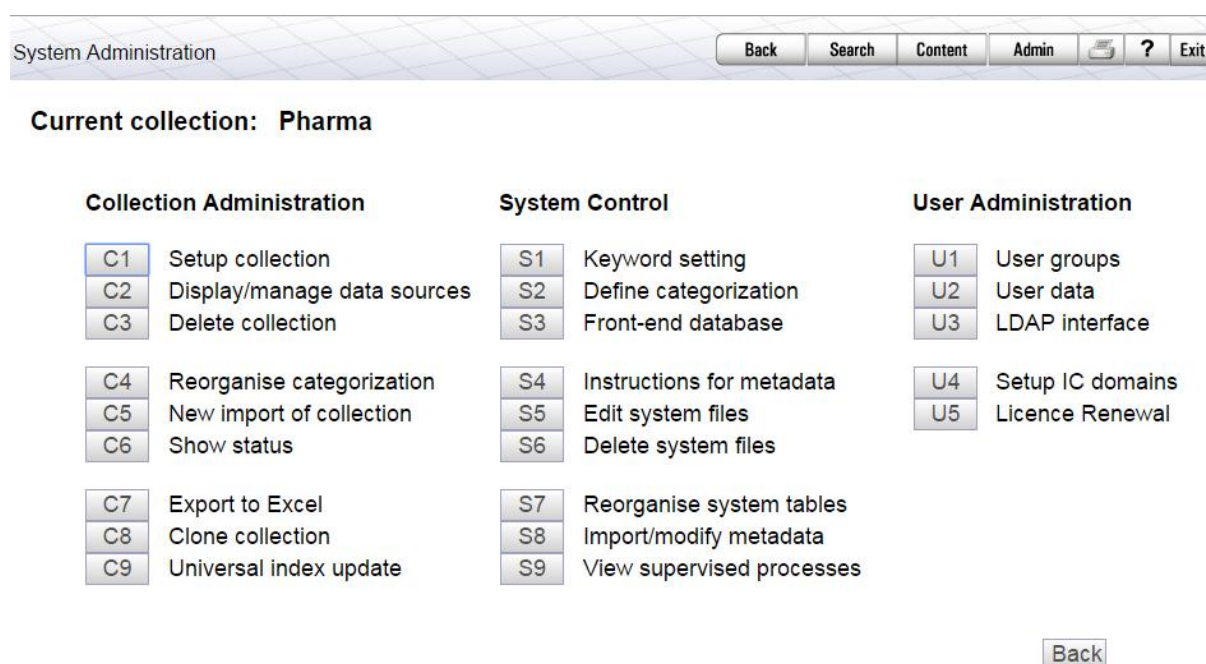
Fig. 12: Menu "S3 Front- End database" in the Admin area

A front end database must be given a name and resides in the central InfoCodex program directory (e.g. c:\infocodex) and is available to all IC domains

Allocation of a Front End Database to an InfoCodex Collection

To allocate front end database it must be explicitly assigned at the time of collection creation.

See: Admin → Set up / delete collection → entry in the field "Front-end linguistic database"

Fig. 13: Collection settings:: Front end database

If a front end database is assigned to an existing collection, the collection's status is set to –99 (i.e. the collection must be freshly generated, see Section 6.2).

5.3   Predefined Categories / Structures

InfoCodex has the ability to identify the main topics in a document collection with no human intervention. Sometimes, however, a user might wish to classify documents in predefined categories of his choice, even if this may lower the overall structurability of the collection. Typically, this is the case for process automation tasks, in which the system has to reproduce organizational structures, or for exploratory tasks, in which a user wishes to home in on a particular subject.

InfoCodex offers three possibilities to apply predefined categories (specifications for the arrangement into the information map):

●	Ad-hoc categorisation	Simple fashioning of the categorisation by mouse clicks
●	Analytical categorisation	Effective, although relative complex specification of the arrangement of the categorisation
●	Fixed, trained categorisation	Fixed specification of the main themes and training of the categorisation using sample collections

In the third case, the categorisation is "*trained*" and remains as a fixed specification for the document allocation of new collections. (i.e. the neural network will not be used in the last categorisation phase). See below for a short overview of how the categorisation works to assist with the understanding the three different procedures.
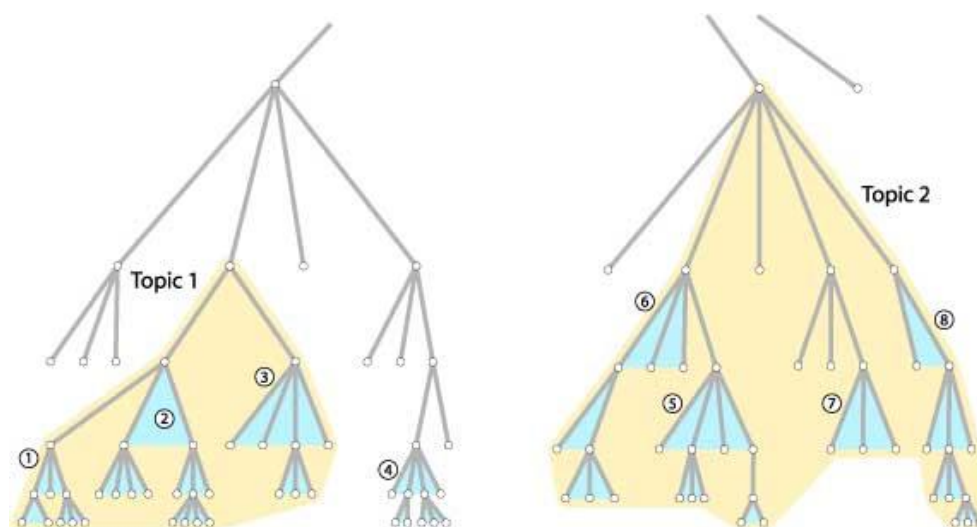
Combining Agglomeration Clusters into Thematic Topics

Fig. 14: Combining Agglomeration Clusters into Thematic Topics

InfoCodex is equipped with a universal taxonomy. In order to be able to optimally categorise the documents in a given collection using a self-organizing neural network, a 100-dimensional content space is tailored to a given collection. This involves determining up to 98 component structures from the taxonomy tree which most reflect the terms in the entire collection (①,②,③ etc. as above). Neighbouring component structures are amalgamated into the main themes (Topic 1, Topic 2 etc. as above). The 98 component structures, together

with a component for the graphical and numerical content ratios and a residual component, form the coordinates of the 100-dimensional content space.

These 98 characteristic components have, of course, a very significant influence on the categorisation by the neural network. The influence on the categorisation is derived, at least partially, from the specification of the characteristic components.

To view the automatically determined components for a given collection, use menu point: → **Admin** → **S2 Define categorisation** → V**iew characteristic components** (see figure below).

Characteristic components for the document collection of "Pharma"

	component	level	relevance	main topic
1.01	**human activity/motif**	6		
	human activity/motif	6	981	
1.02	**psychological feature**	6		
	psychological feature	6	1081	
1.03	**psychological state**	6		
	psychological state	6	821	
1.04	**physiological state**	6		
	physiological state	6	1362	physiological state
	disorder	5	614	physiological state
	illness	5	1210	physiological state
	disease	4	1317	physiological state
	infectious disease	2	446	physiological state
	tumor	3	362	physiological state
	infection	5	384	physiological state
	injury	5	295	physiological state
1.09	**trouble**	6		
	trouble	6	912	
2.01	**social group**	6		
	organization	5	384	
2.04	**administrative district**	6		
	country	5	702	administrative district
	European country	4	598	administrative district
	city/town	5	410	administrative district
2.07	**legal system**	6		
	legal system	6	1480	
3.02	**economic system**	6		
	economic system	6	1015	
3.03	**business organization**	6		
	management	5	587	
	place of business	5	294	

Fig. 15: InfoCodex' automatically generated conceptual regions

This list shows, in particular, which concepts, and with which relevance, are present in the given document collection.

Method 1: Ad-hoc Categorisation

This simple, clear and mouse-click operated categorsation method is available under

→ **Admin** → **S2 Define categorisation** → **Ad-hoc categorisation**.

After giving the maximum number of nodes to display (recommended: 300 - 1000) the system displays those parts of the taxonomy tree that are relevant to the current collection. The highest hierarchy levels are hyperlinks and the detail to deeper levels are graded to the right. The grey shades of the individual fields represent the relevance in the given collection: dark

grey fields being more relevant than light grey fields. The coloured areas represent the areas of the taxonomy tree which are used as main topics in the collection.

SCIENCE/ ART/ DISCIPLINE				
science/research	linguistics			
	mathematics	applied mathematics		
	natural science	biology		
		chemistry		
		earth science	geology	
		medical science	pharmacology	
			medical term	
		physics	mechanics	
			nuclear physics	radiology
			particle	elementary particle
	psychology			
	research			
scientist	biologist			
	chemist			
art	music			
special discipline				
investigation	analysis			
	determination	designation	diagnosis	
	examination			
	information	information gathering		
		news		

Save   Cancel

Fig. 16: Representation of an ad-hoc categorisation

In order to influence the classification criteria of the system it is sufficient to select, delete and merge the desired taxonomy nodes to form the desired main subjects. The following operations are possible:

1. Clicking a node within the grey area: the branch of the taxonomy tree that begins at the clicked node and encompasses all other nodes from the corresponding lower hierarchical levels is assigned a colour and corresponds henceforth to one of the main topics. The name of this topic may be edited.

2. Clicking a node within a colored area: the branch that begins at the clicked node is marked and it can
   a) either be deactivated (i.e. it will become a neutral grey zone)
   b) or be declared as a new topic
   c) or be merged with another topic

Once all desired modifications have been made, the configuration can be saved as a predefined categorization by clicking "save". The user is prompted for the table name which must have the extension ".pm" (predefinded map). The "pm"-table created is available to all collections in the current IC domain.

To bring the ad-hoc categorisation into effect, use the the function **C4 Reorganise categorisation**.

Method 2: Analytic categorisation
This theoretically, well founded method consists of the direct assignment of the abovementioned 98 characteristic components (dimensions of the content space). It is evoked with → **Admin** → **S2 Define categorisation** →   **Analytic categorisation**.

Step 1: Load existing component selection
Selection using the automatically generated component list as a starting point.

Step 2: Modify the component selection

Check the desired nodes in the taxonomy tree that are required to be component characteristics (see figure below). By selecting a node, all subsequent underlying nodes (hierarchy levels) are also included (e.g. 104.05 upto and including 104.0511). If, however, an additional node is checked in an underlying level, this will be excluded from its "parent" component and become one in its own right (in the example below, 104.08 comprises all underlying nodes except for 104.0801, which forms its own component).
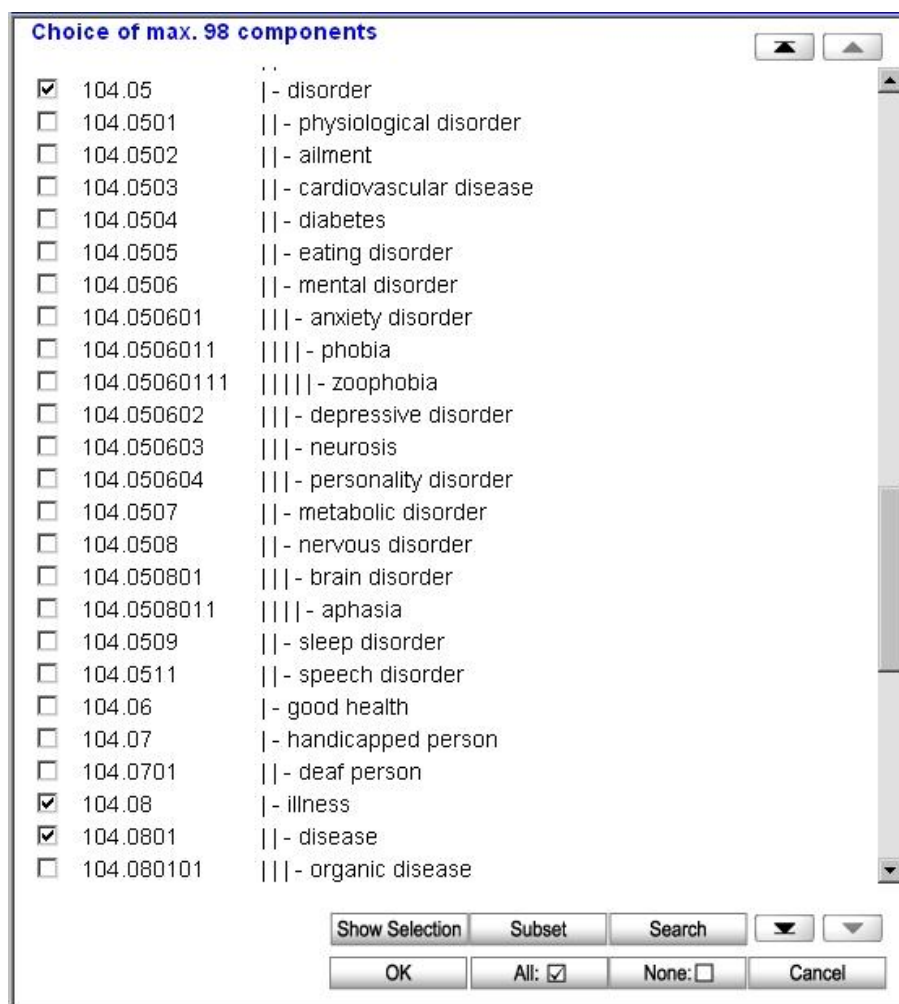
Fig. 17: Selecting components

If fewer than 98 components are selected, InfoCodex will automatically supplement the selection to 98.

Step 3: Defining main topics

Select 2 to 20 main topics into which the information map is to be divided.

Step 4: Assignment to main topics

Assign the 98 characteristic components to the main topics defined in step 3.

After step 4, the categorisation criteria are saved in the PM table, and a reorganisation of the information map should be performed.

Method 3: Fixed, trained categorisation

With this variant, the categorisation is "trained" with the help of a sample set of documents of familiar content. The documents in the sample set are then assigned to the particular main topics of the information map.

The categorisation, in this case, is predetermined in the training phase. It remains fixed for all subsequent imports and the automatism of the neural network is deactivated.

The general procedure is as follows:

Step 1: Prepare a sample set for the map training

Define a list of 2 to 20 main topics corresponding to the target structure of the information map, e.g. "monopoly", "tax" and "state property". Then, prepare a sample collection with familiar content such that the documents characterising specific topics are located in separate directories (one directory per main topic, each containing at least 20 documents).

Step 2: Create a script file for the map training

The instructions for building the predefined categorisation from the prepared sample set are entered into a script file located in the root directory of the current IC domain e.g.

```
"c:\icdata\instruct.txt"
```
where "instruct.txt" is an arbitrary name for the script.

The script should have the following format:

```
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
```

The following keywords are used:

Topic   The topic names must be written in English and they must correspond to InfoCodex hypernyms (nodes in the taxonomy tree). The hypernyms from user-defined front end databases are also permitted.

Dir   The documents of this directory and its sub-directories should be representative of the corresponding topic. These documents, together with the optional keywords, are used for the training of the map.

KW   Optional keywords

      KW=wine, liquor, cigars   means:
add the terms "wine, liquor, cigars" to each document of the corresponding directory (only in the training phase)

KW=*IDB: state property   means:
add all terms from the InfoCodex linguistic database belonging to the hypernym "state property " to each document in the corresponding directory (only in the training phase)
(analogously for the front end database with KW=*UDB: …)

The KW specifications are not mandatory. They can be used for a supplementary characterisation of the target topics.

Step 3: Network training using the sample collection

Create a new collection (by clicking the corresponding symbol on the left side of the screen), enter the collection name and choose "Advanced options"..

Create New Document Collection

Collection name   SOGEI - fixed predefinded map

[Advanced options]

Fig. 18: Create a new collection

Enter the name of the script file, preceded by an asterisk, in the field "Predefined categories", e.g. *instruct.txt. It is recommended to prescribe the number of columns of the map, e.g. 20. ; otherwise, the dimension of the generated neuron matrix could turn out relatively small if the sample collection is not large enough.

Optional classification setups

Classification

Keyword table	
Predefined categories	*instruct.txt
Metadata instructions	
Frontend linguistic database	
Indexing	☑ incl. **indexing of** exact terms
Special options	☑ Generation of abstracts ☑ Build document families
Import mode	☑ for identical filenames → update old entry

Map formation

No. of map columns	20	(no. of col. = no. of rows)
No. of main topics		(coarse subdivision of the map)

[OK save]   [Cancel]

Fig. 19: Advanced Options: pre-defined categories and the dimension of the information map

The map training commences immediately after clicking the button "**OK save**".

Step 4: Optional improvement of the generated map

After the automatic generation of the map, the ad-hoc categorisation can be used for adapting the arrangement to one's own particular perspectives. Afterwards, execute the function "**C4 Reorganise categorisation**" to make the improvements effective.

Step 5: Reset status and add documents

The predefined map is now set up, and you can load any document collections into the fixed network along with their predefinded categories. The structure of the network remains unchanged.

Prior to the import of the documents, the status of the trained map has to be reset:

Set up/delete Collection

collection-ID: 10   status: ▼ ,?
collection name
german: SOGEI - predefined map
english:

Select from 4 positions	- Micro... _ □ ✕

Collection status

  Status = -1
0   Reset to "no documents imported"
-1 OK, document import completed
  >0 import processing

[Cancel]

Classification
Keyword table
Predefined categories   ▼ *instruct.txt+C:\domewae\publicfinances

Fig. 20: Setting the collection status to " no documents imported"

After resetting the status, the import can take place with the newly "trained" map.

6. Auxilliary Functions

6.1 Excel Spreadsheet Import / Export

 InfoCodex offers several possibilities for the exchange of server data with Excel spreadsheets on the client. For this transfer to be possible, one must download and correctly install the "plugin" available on the InfoCodex start page (see Section 4.2 Part 1 #1).

 Suggested uses are the output of search results to Excel spreadsheets (see Section 3.2 Part 1) and the output of categorisation results and keyword setting (see Section 5 Part 1).

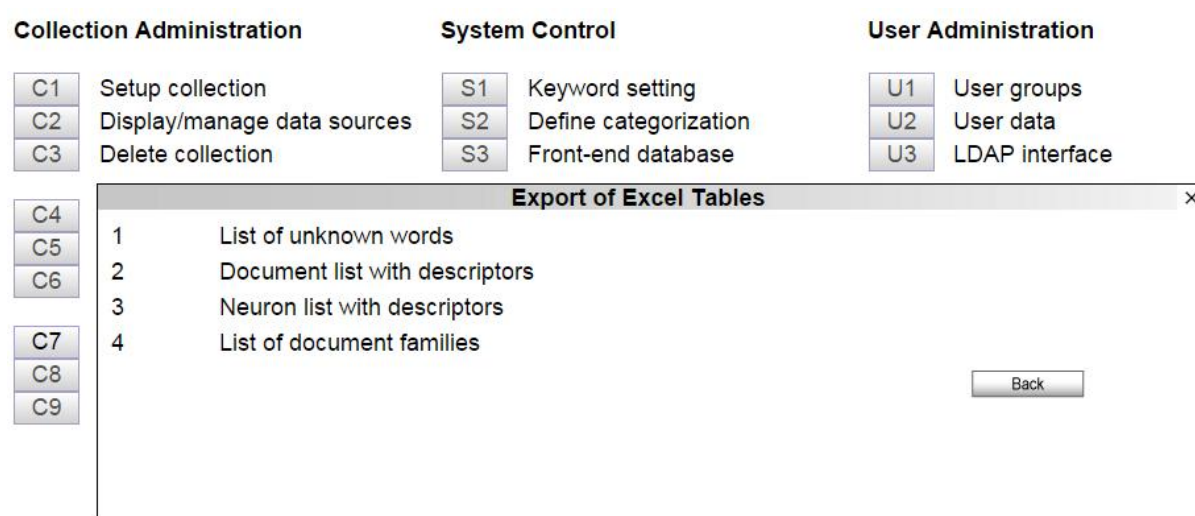Further possibilities exist under: **Admin → C7 Export to Excel**

Collection Administration

C1	Setup collection
C2	Display/manage data sources
C3	Delete collection

System Control

S1	Keyword setting
S2	Define categorization
S3	Front-end database

User Administration

U1	User groups
U2	User data
U3	LDAP interface

C4		**Export of Excel Tables**	✕
C5	1	List of unknown words	
C6	2	Document list with descriptors	
	3	Neuron list with descriptors	
C7	4	List of document families	
C8			Back
C9			

Fig. 21: Export to Excel

A noteworthy feature is the output of words occurring in a collection that are unknown to the InfoCodex database. This list can be used, on the one hand, to locate orthographic errors in the documents; on the other hand, it can be used as a basis for the construction of a front end database.

6.2 Regenerating Collections / Updating Hit Statistics

These functions are used mostly after an update of the InfoCodex linguistic database or when a front end database has been assigned to a pre-existing collection. Such changes have the consequence that the coded extract of the documents is incorrect and must be reconstructed.

This allows the regeneration of a collection without the need to repeat the previously performed import instructions, keyword table construction and categorisation.

Hit Statistics

This is a continual tracking statistic on the number of clicks in the individual neurons i.e. the viewing of documents per user. It can be used to locate people who often consult certain document groups, and who, therefore, could possibly be experts on a certain subject.

The hit statistics remain in tact after a simple reorganisation. If a regeneration is selected (with a new import of the documents), the user is given the option of keeping the existing hit statistics or resetting them.
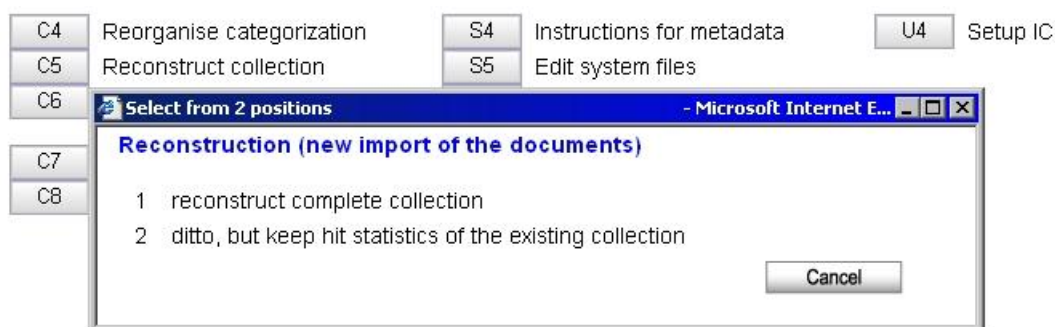
Fig. 22: Regenerate a collection, with or without updating the hit statistics

6.3 Copying / Editing / Deleting

Copy function: **C8 Copy collection**

Copy an existing collection to a new collection ("Cloning")

Edit script files: **S4 Edit system files**

Change system settings by the system administrator ("options.ictxt" etc.; see Section 7).

Delete **S4 Delete system files**

Deletion of keyword tables, PM tables (predefined categories), front end databases etc.

6.4 Job Scheduling

If an import of documents is to be periodically repeated, the import can be set up as a batch job for recurrent processing (see Section 4.3 Part 1):
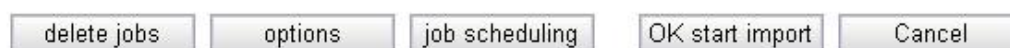
Under "**Job Scheduling**", a dialog appears for the specification of the start time and periodicity of the processing.

Fig. 23: Options for periodic imports

Under "Actions to perform" one selects whether a collection should be newly generated (= "Initial Load") or if the documents are to be added to the existing collection (= "Incremental Load").
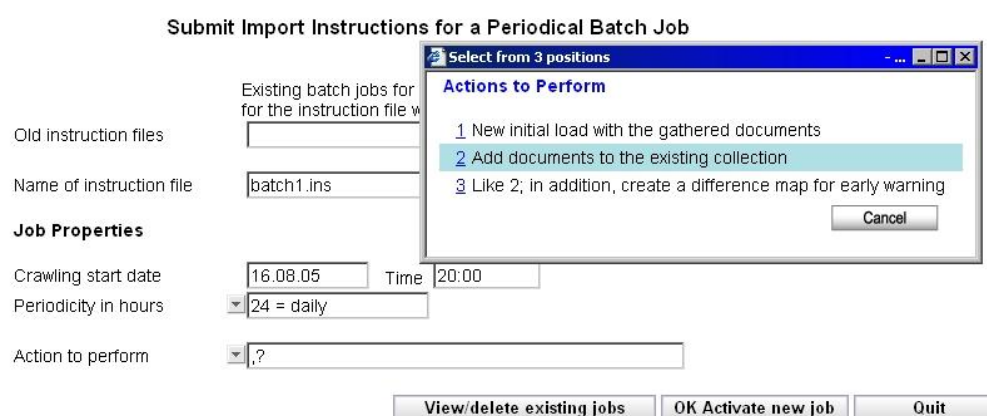
By repeated batch job imports for documents having the same file name as those already in the collection, one has the option of either appending them to the collection or updating their existing entries. This option can be set when setting up the collection using the check box for "Import mode"

If this option is not activated, the documents will be added again by each successive import (see. Chapter 4.3 from Part 1 of the user manual).

6.5 Monitoring and Permanent Background Processes

The following InfoCodex processes run contiually in the background:

Wimonsrv.exe

This program is installed as a system service on a Windows machine. The service starts and and monitors wimon.exe and processes user logins in the case that the InfoCodex user administration is bound to a Windows domain.

Wimon.exe

The InfoCodex scheduler is started from wimonsrv.exe on a Windows platform and as a "crontab" entry on Unix/Linux platforms.

This program is called every minute and checks the status of the communication daemon wihjxd.exe; the WIDAS/HTML-based administration interface and the program icapidaem.exe which coordinates the InfoCodex API. It also administers, every 30 minutes, any import jobs that must be started.

ICRegen.exe

This program is started periodically from wimon.exe to execute any batch jobs that are due.

It reads the file "importschedule.ictxt" in the central InfoCodex software directory and carries out the appropriate instructions. It also updates the file "importqueue.ictxt" with the jobs that are set up or deleted by the users.

The periodicity of its execution is given as a number of minutes in the first line in the file "importnext.ictxt".

ICWatch.exe

The InfoCodex watch dog, which checks and corrects any program crashes that may have happened, e.g. killing and restarting hanging processes, or restarting crashed processes at the correct phase of their execution.

ICDelete.exe

The uncoordinated deletion of large masses of data from disk (e.g. a collection of several million documents) can substantially debilitate performance and cause the user unacceptable waiting time. This program makes sure that such operations occur in the background when the server is otherwise not so busy.

7. System Setting

7.1 Proxyserver (proxy.ini)

When the InfoCodex server must access the Internet or the intranet by means of a Proxy Server, the proxy address has to be specified in the file "proxy.ini" in the InfoCodex program directory (e.g. in "C:\infocodex"). The following items must be specified:

1	`www-proxy.intern.local`	Name or IP address of the proxy
2	`8080`	Port of the proxy
3	`Infocodex`	Username (optional)
4	`Secretpwd`	Password (optional)

If no "proxy.ini" is present, InfoCodex uses the proxy settings in the browser installed on the InfoCodex server.

7.2 Processing options (options.ictxt)

The InfoCodex background processes can be controlled with several different parameters. They concern mostly the spider agents and the import of documents, the text mining and the categorisation and indexing processes.

The parameters are also stored in a central .TXT file "options.ictxt" in the InfoCodex program directory (case insensitive)

Format of options.ictxt:

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

The default values for these parameters are outlined below:

SECURITY = 0

Specification of the security level (see. chapter File System Security)

RETARDANT = 10

The import and the analysis of document collections are performed by background jobs which should not terminate before the whole job is completed. Some Windows versions show some problems with their disk cache management when extensive output operations are involved. This might result in a job termination before orderly completion.

In those cases, you should make use of the RETARDANT option with the effect that the background job will make controlled pauses at some critical steps.

Recommended range: 0 for well conditioned systems
10 for Windows systems with very high clock frequencies or with 2 or more processors
30 for extreme situations

CLEAN = 30000

The text mining step needs large memory resources. As a precaution, InfoCodex makes a controlled garbage collection (memory clean) when the number of newly recognized words exceeds a given limit.

Recommended range:	30000	for systems with 256-512 MB RAM (=default)
	15000	for systems with 128 MB RAM
	50000	for systems with more than 1 GB RAM

NUMBER = 20

Documents containing large statistical tables would lead to enormous sets of different synonyms for all the encountered numbers. To prevent a very large overhead, the maximum number of different numbers to be taken into account within a document can be restricted by the NUMBER option.

| Examples: | 20 | if only the first 20 numbers in a document are of interest (Customer no., Ref.no. etc.) |
| | 1000 | if the first 1000 numbers are of interest |

TRACE = 0

When documents are imported from an Internet search engine or an Internet / intranet website and when InfoCodex does not find all links that it is supposed to find, it is recommended to use the TRACE option.

| Examples: | 0 | no special link tracing |
| | -2 | record all recognized links in the logfile "import.log" |

SHOWMSG = 0 (Microsoft Outlook)

When e-mails are imported, their content is saved in a text file to be displayed later. Messages imported from Microsoft Outlook are also saved in the proprietary MSG format. MSG files appear on the client just like any other e-mail message, including headers and attachments, but require a compatible version of Outlook installed on the client.

| Possible values: | 0 | Show plain text file |
| | 1 | Show full message, if available (else resort to plain text file) |

OL_LocalPrf (Microsoft Outlook)

Outlook profile for Outlook installed on the InfoCodex server used to import messages from PST files (Outlook archives).

| Example: | OL_LocalPrf = infocodex-lokal |

OL_ExchangePrf (Microsoft Outlook)

The profile to be used for Outlook that runs on the server where InfoCodex is installed, and to be used for the import of mail from Exchange Server.

| Example: | OL_ExchangePrf = infocodex |

ExchangePwd (Microsoft Outlook)

Password for the login to the Exchange server by the profile "OL_ExchangePrf" (if a password is really required for "OL_ExchangePrf").

OL_CleanPST = 0 (Microsoft Outlook)

Instructs InfoCodex to close all open message stores (PST files) before the import of a PST file.

Possible values: 0   Ignore open message stores

1   Close all open message stores (except for the default one)

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

When an email import should be performed from the client side Outlook mailbox, the imported emails must be temporarily copied to the InfoCodex server. These are saved to temporary subdirectories under "ClientShare" specified network directories.

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

This is the local name of the directories on the InfoCodex server mentioned in ClientShare above.

TranslateToEN=0

Flag for the automatic translation to English of documents in languages other than the standard InfoCodex supported languages (German, English, French, Italian or Spanish).

Possible values: 0   do not translate
1   use online translation tool (Google translate or Systran) to translate documents into English.
Primarily supported languages: Russian, Chinese (traditional or simplified), Arabic, Indian (Urdu, Hindi, Devangari), Japanese, Portuguese.

N.B. This option must be set **before** a new collection is created to have effect.

AccessLog=0

Logging of successful and failed login attempts.

Possible values: 0   No logging

1   All login attempts are logged into a logfile named **icaccess.log**

CompressFiles=0

The generated text files and pseudo HTML files can optionally be compressed. The compression saves disk space and may increase performance.

This option has no effect on non-NTFS volumes or operating systems other than Windows.

Possible values: 0 No compression
1 Generated TXT-files and pseudo HTML files will be compressed

CutUArgs

A space delimited list of variables to be removed from URLs.
Some websites insert session IDs or other volatile information into hyperlinks, as in
"http://www.example.com/page.php?SESSIONID=12345678". To InfoCodex this means that
even when the documents are still the same in subsequent imports, the URLs will change
every time.
InfoCodex will work around this by removing any variables from a URL that match any of the
regex fragments in CutUArgs, e.g.:

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

MailServer, MailPort, MailFrom, AlertsFrom, MailAuth

In order to send news alerts and other notifications, the following three options need to be
configured:

MailServer   The host name or IP address of the mail server to use.
MailPort   The port to use (typically 25).
MailFrom   The sender, e.g. "InfoCodex Server <infocodex@example.com>".
MailAuth   The authentication (if necessary), e.g. "MailAuth = myUser:myPwd".

InfoCodex will only attempt to send out e-mail if the first three options are present.

If the server is used to send alerts by mail, the following option is needed:

AlertsFrom   The sender, e.g. "InfoCodex Alerts <alerts@example.com>".

TxtFiles = txt

A space delimited list of file extensions to be considered plain text (ASCII) files during import.
e.g.:

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

7.3   Interface to Lotus Notes

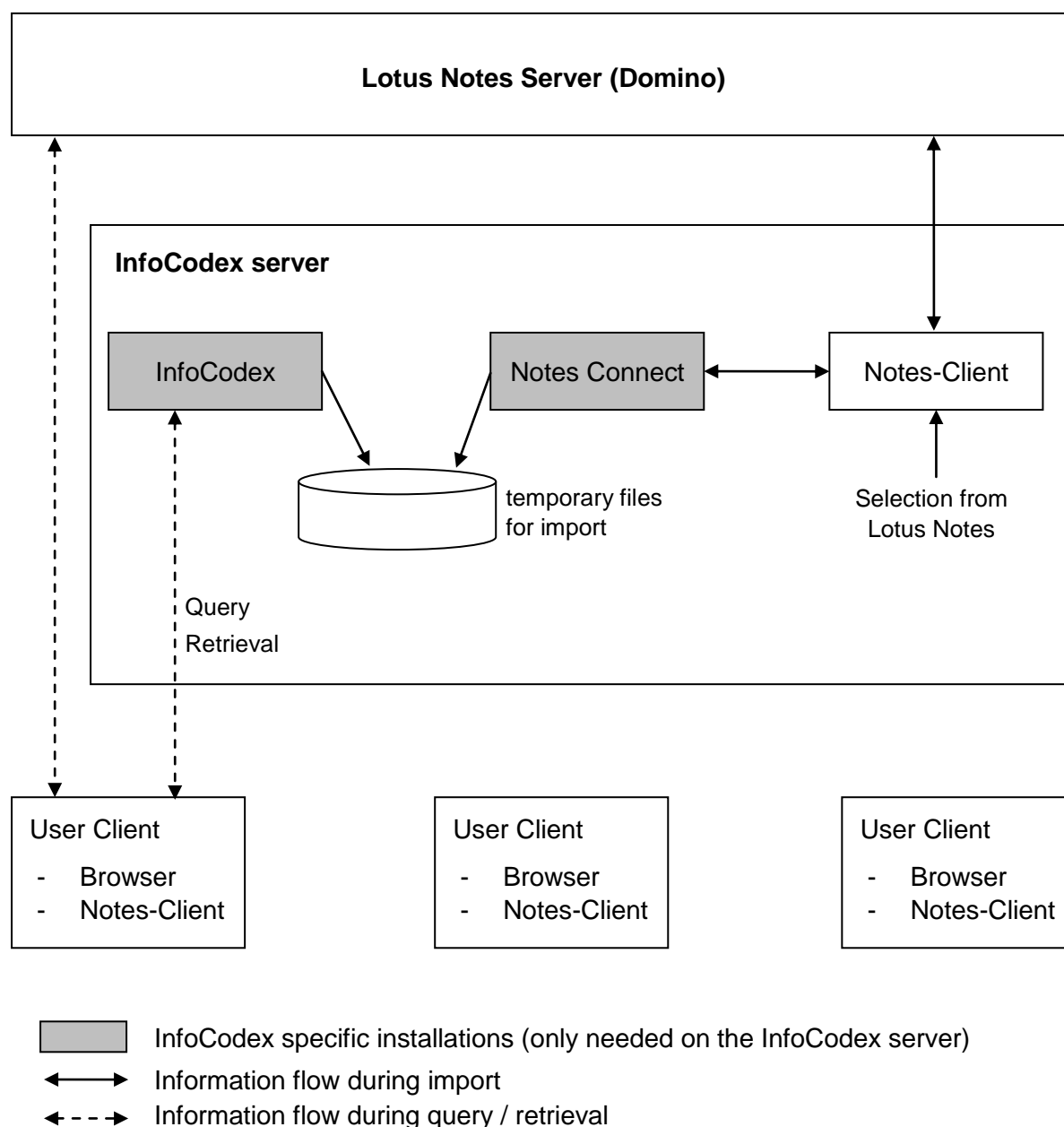The special InfoCodex module "NotesConnect" and a Notes Client must be installed on the server.

Fig. 24: Components and Information Flow at Import from Lotus Notes

Importing Notes documents

The selection of Notes documents (and their attachments) from the various Notes databases occurs with the Notes client (this must possess the necessary privileges).
The selected documents are also written to temporary files by a background process and deleted by InfoCodex after the analysis phase.

The Notes documents and attachments are now available to InfoCodex for query and retrieval. The links to the Notes databases are stored in InfoCodex so that the display of the individual Notes documents can be accessed via Notes Client.

Access Rights

The Notes access rights to imported Notes documents can also be recognized and user access coordinated via LDAP.

7.4 Interface to Microsoft Outlook and Exchange Server

These interfaces require that Outlook is installed on the InfoCodex server and the correct entries present in "options.ictxt" (see Section 7.2).

7.5 Restricting File Types, Third Party Conversion Programs

Normally, InfoCodex determines the file format based on the content and ignores the fi,e name. In this way, documents whose file extension is inconsistent with their content can be correctly recognised and imported (e.g. RTF disguised as .doc, .xls/.xlsx/.csv etc.). However, with very large amounts of incorrectly named files can significantly slow down the import process. In such circumstances, the import can be expedited by restricting the range of file extensions to be taken into account. For such cases, one can create a file called "extensions.ictxt" either in the central InfoCodex directory (for the global case) or in the independent collection directory (valid only for that particular collection).

If such an "extensions.ictxt" file is to be used for a specific collection, it must manually entered in the batch-job instruction file (e.g. Batch1.ins) using the keyword EXTENSIONS and explicitly defining the path to that file e.g.: "EXTENSIONS=c:\xxx\myextension.ictxt".

Example of a "extensions.ictxt" file:

```
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
```

Third Party Conversion Programs

It is also possible to assign external, third party, text-conversion programs for the indexation for file formats that are not supported by the standard InfoCodex installation. The allocation of a third party program to a specific file extension can be assigned in the "extensions.ictxt" file as in the following example:

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

The parameters symbolised by <inputFile> and <outputFile> are obligatory. In the case where a text conversion program is able to extract descriptive metatdata, the output file for the metadata is given in the optional parameter <metadataFile>. The syntax and keywords should have this form:

```
AUTHOR: InfoCodex
SUBJECT: Targetted Market Observation
TITLE: Early Identification of Market Trends and Facts
MODIFIED: 12.04.2008
```

Important Notice:

When using third party conversion programs,

- either all the file extensions of the files to be imported have to be registered in "extensions.ictxt" (doc, docx, xls, xlsx, pdf etc.)

- or the character "*" is to be inserted in "extensions.ictxt", which means that all standard extensions are to be taken into account.

Example 1:
```
    *
htm,html=c:\xxx\myhtmlconv <in> > <out>
```

Example 2:
```
htm,html=c:\xxx\myhtmlconv <in> > <out>
```

In example 2, *only* the files with extensions ".htm" and ".html" will be interpreted.

Import using iFilter

Under Windows, an iFilter is a program that can extract text from a document of a specific format. It is usually available with the manufacturer's software. InfoCodex can be instructed to use such iFilters by means of the Windows utility "filtdump.exe" (delivered with InfoCodex).

If an iFilter is installed and is to be used for the InfoCodex indexation, it can be specified in "extensions.ictxt" as follows:

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

where the example "dwg" would be the typical file extension of that file type.

7.6 Drive Mapping ("drives.ictxt")

For the case where users in a Windows network environment have allocated special names for network drives, these mappings may be specified in a TXT file "drives.ictxt" in the central InfoCodex program directory. This enables users to see their familiar drive labels like "R:", "S:" etc. instead of the absolute network drive names.

Example of a "drives.ictxt" file :

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

The convention "<user>" means that the user's username will be substituted in the network path, e.g.

"\\fal008s\private\smith" for user "smith".

7.7 Extended Daemon Settings ("icmonitor.ictxt")

This file contains the parameters for the monitoring program "wimonwin" (for Windows) or "wimonux" (for Linux and Unix), respectively, and the communication daemon "wihjxd". It is generated automatically at the installation of InfoCodex (see chapter.Monitoring).

Special entries:

```
Cleanup 03:00
```

The program "wimon" performs a daily cleanup function of deletion of leftover temporary files and a restart of the communication software. This parameter stipulates the time of day that it should be executed.

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

The above entries are required for the execution of the HTML environment and the InfoCodex API. Normally, they should not need to be changed.

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

The keyword DAEMON refers to processes that need to be monitored and, if necessary, restarted by the process "wimon".

7.8 Configuring the InfoCodex API Daemon (webserver.ictxt)

The file "webserver.ictxt" contains the server names relevant to the clients and the root directory of the web-server. e.g..:

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

7.9 Pausing large import processes

In order to prevent resource-intensive imports from slowing down small ad-hoc imports and interactive user processes, import processes for big collections can be configured to pause at certain times.

Import times can be configured by creating a file **pause.ictxt** in the collection directory, using the following format:

- One line per time range

- Lines starting with "#" or ";" are ignored

- Days of the week, begin and end time, e.g:

```
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)
```

An example of a pause.ictxt file can be found in the InfoCodex program directory. This file is solely for demonstration purposes and has no effect.

8. Licence Management

The client's licence number must be entered at installation time. The validity of the licence is checked periodically. If the validity is established, the InfoCodex installation is activated for a few weeks or months. A new test occurs when this period runs out.

8.1 Automatic Activation

Normally, an activation is valid for several weeks. InfoCodex automatically tries to extend the activation. An external HTTP connection to the licence server *lic.infocodex.com* must be enabled for this to function.

If the automatic licence extension fails, several days before the expiry date, the following warning will be displayed.
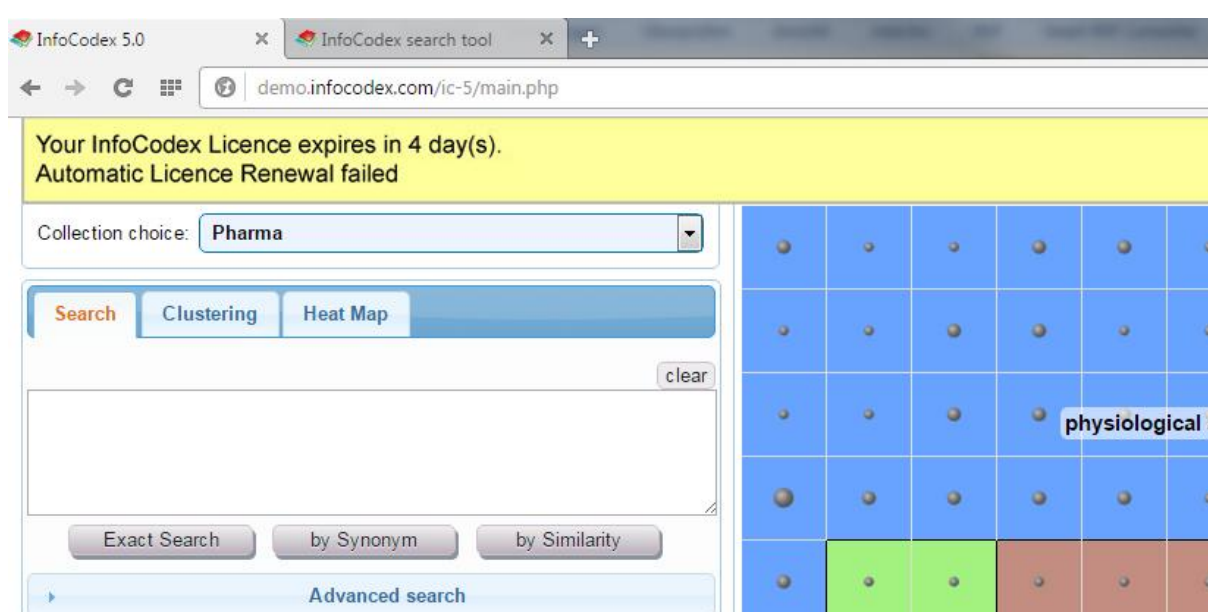
Fig. 25: Warning message by expiring licence

8.2 Licence Renewal / Clearance without Internet Connection

In the case of an invalid licence or a clearance is required and the licence server is unreachable (e.g. when the InfoCodex server has no internet connection), the following error message will be displayed.
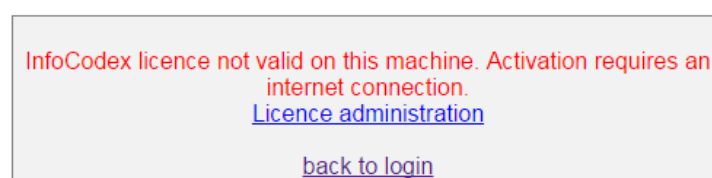
Fig. 26: No valid licence number available.

To renew or activate a licence without an internet connection, select the menu points

"**Admin**" → "System administration" → "**U5** - Licence Renewal", and you will get the following instructions for a manual licence clearance:

Fig. 27: Manual licence clearance

Copy the displayed text (from -------- to --------) and send it per email to support@infocodex.com. You will receive a clearance code that you can enter under "OK – Accept the received e-mail text". To which the InfoCodex Server will be validated.