SIARDexcerpt Manual

Content

1 Preface

SIARDexcerpt is a Java-based application that searches and extracts individual records of SIARD files[1], then converts them into a human readable form using a user-specific or a generic stylesheet. It is an open source application under a GPL v3+ licence. SIARDexcerpt uses unmodified components of other manufacturers by embedding them directly into the source code. Users of SIARDexcerpt are requested to adhere to these components' terms of licence. Please refer to chapter 8 for further information.

SIARDexcerpt complies with the following requirements. See also chapter 9.2 for a diagram of the functional principle.

Initialisation: During initialisation the SIARD file is unpacked into the working directory and the desired configuration is copied to the pre-defined location. If required, the configuration is completed automatically according to metadata.xml as described in chapter 9.1 and temporally saved as SIARDexcerpt.conf.xml.

Search: After initialisation the matching lines are searched using grep. The asterisk (*) serves as a wild-card character. SIARDexcerpt copies the matching lines and outputs twelve pre-defined columns as a preview. A stylesheet permits the display of the result in Internet Explorer. The search result is saved into the configured output folder.

Extraction: The extraction can start once the primary key is known. The extracted result is saved into the configured output folder. A stylesheet permits the display of the result in Internet Explorer.

End: At the end, the temporary configuration file SIARDexcerpt.conf.xml and the unpacked SIARD file are deleted.

The results (including information on inconsistencies and errors) are output for every step and written into a validation log.

2 System requirements

Microsoft Windows 98 or later

128 MB RAM or more (512 MB or more is recommended)

20 GB disk space or more

Java Runtime Environment (JRE) Version 6 (resp. 1.6)

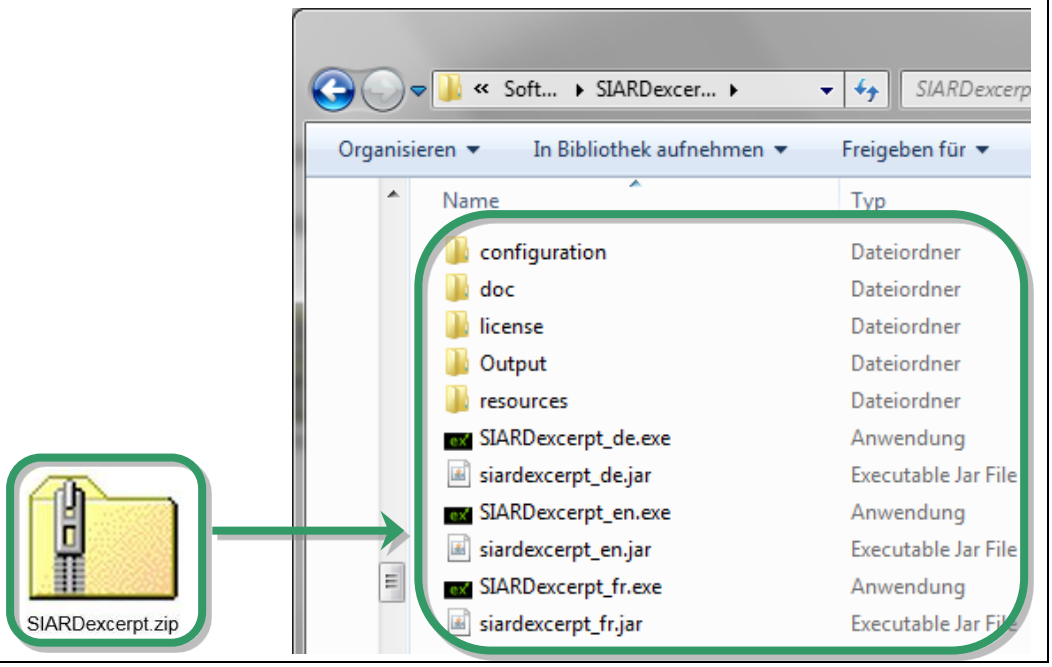
3 Open issues / Feedback

Open issues ranging including bugs, requested features, and questions, are listed on the software development platform GitHub at https://github.com/KOST-CECO/SIARDexcerpt/issues and can also be communicated to kost-val@kost-ceco.ch.

These issues are managed by the development team. Any and all contributions are welcome.

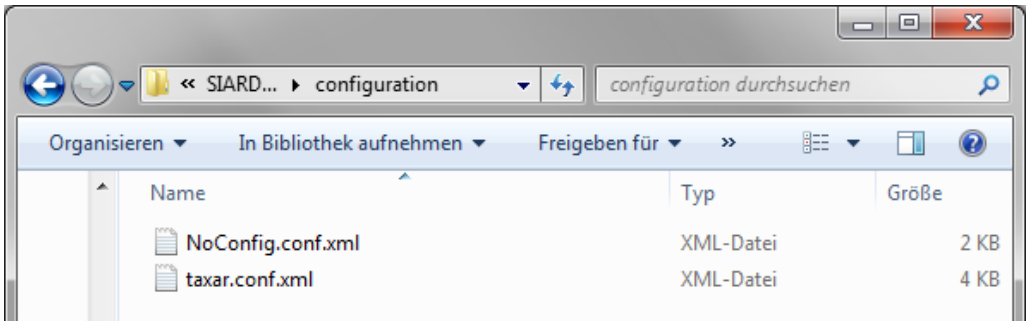
[1] The SIARD specification (*Software Independent Archiving of Relational Databases*) can be downloaded from the eCH website: http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0165.

4 Installation

1	Download SIARDexcerpt and unpack the ZIP file into the relevant folder. 

5 Configuration of SIARDexcerpt, of SIARD search, and extraction

Some configurations of SIARDexcerpt need to be specified previously in a configuration file. An XML editor or Notepad ++[2] should be used to modify XML files.



The basic configuration file is "NoConfig.conf.xml" in the "configuration" folder. It can be used for all SIARD files. The "taxar.conf.xml" file is an example of a specific configuration file for the extraction of the tax register database.

On initialisation a copy of every configuration file is saved as "SIARDexcerpt.conf.xml" in the "configuration" folder, whereby all "(..)" values are filled with values taken from the SIARD file's metadata.xml file. It is recommended to save this file after initialisation under another name. This enables users to subsequently improve the configuration for the next use of SIARDexcerpt.

[2] Portable Notepad ++ is available from http://portableapps.com/de/apps/development/notepadpp_portable.

5.1 Parts of the configuration file

The configuration file consists of several parts that are described in detail in chapter 9.1.

The pre-installed configuration allows immediate searching in and extraction of any SIARD file.

The following is a short description of the configuration parts.

5.1.1 General

xml-Tag	Description: default value
<pathtooutput>	Path to output folder of SIARDexcerpt: **Output**
<pathtoworkdir>	Path to temporary working directory of SIARDexcerpt: **temp_SIARDexcerpt**
<pathtoxsl>	Path to the extraction stylesheet, e.g.: **(..)**
<pathtoxslsearch>	Path to the search stylesheet: **resources\SIARDexcerptSearch.xsl**
<archive>	Name of the Archives: **Archiv**

5.1.2 maintable part

xml-Tag	Description: default value
<mainfolder>	Number of the table in which the search is to be executed, i.e. the main table for the extraction: **(..)**
<mainname>	Name of the table in which the search is to be executed, i.e. the main table for the extraction: **(..)**
<title>	Title of search results: **(..)**
<primarykeyname>	Name of the primary key: **(..)**
<primarykeycell>	Indication of the cell in which the primary key is stored: **(..)**
<cellname1> … <cellname11>	Names of the additional columns for the output of the search result: **(..)**
<cellnumber1> … < cellnumber11>	Indication of the cells in which the columns for the output of the search result are stored: **(..)**

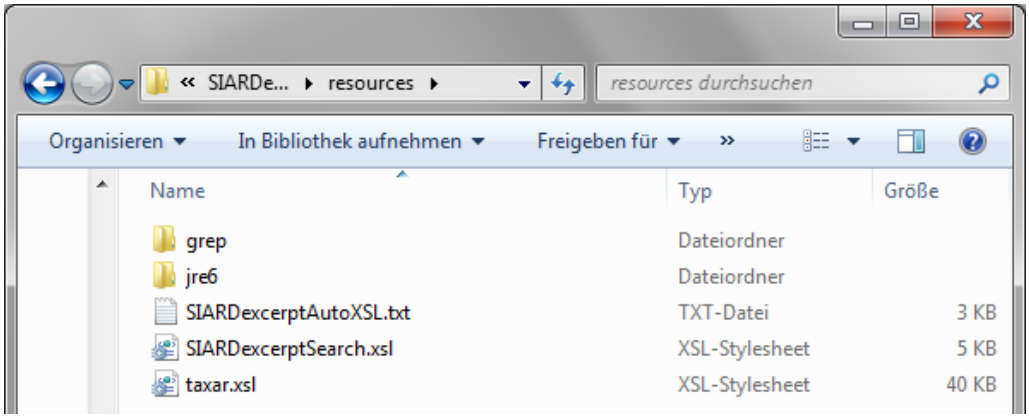
5.1.3 subtables part

xml-Tag	Description: default value
<subtables>	Information[3] about the tables to be joined and the respective keys: **(..)**

[3] For every <subtable>, <name>, <folder>, and <foreignkeycell> are necessary and will be identified.

6 Ressources of SIARDexcerpt

All resources of SIARDexcerpt are stored in the "resources" subfolder.



SIARDexcerpt uses "grep" and "jre6".

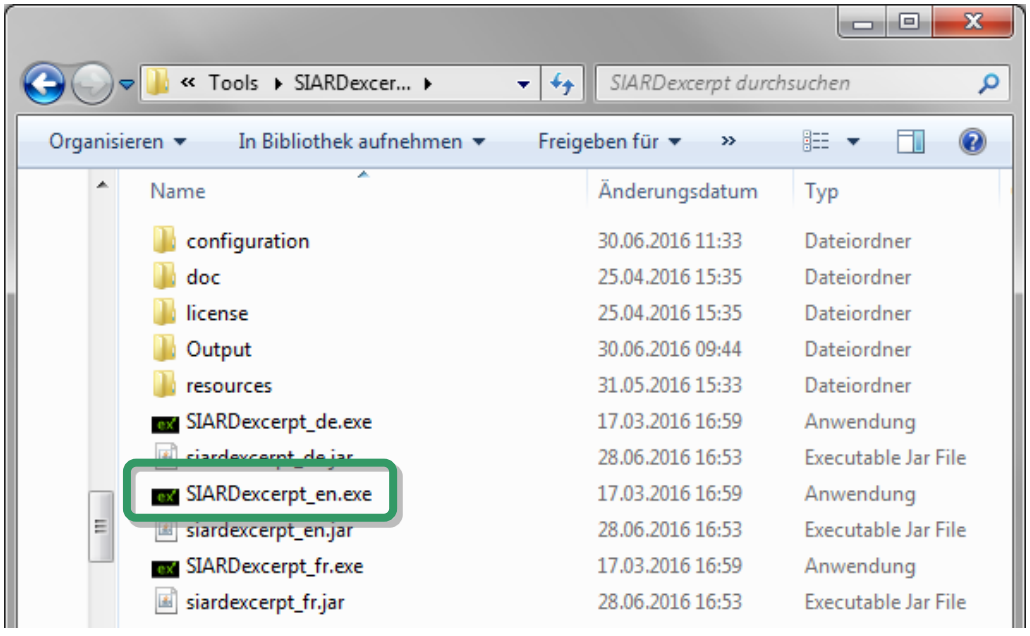
The "SIARDexcerptSearch.xsl" stylesheet allows to display and read the search results table directly in Internet Explorer.

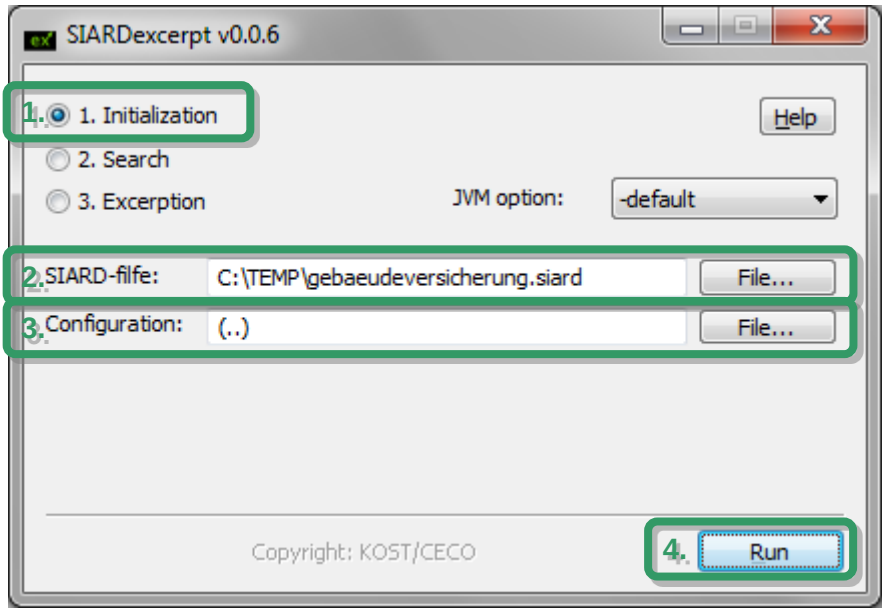
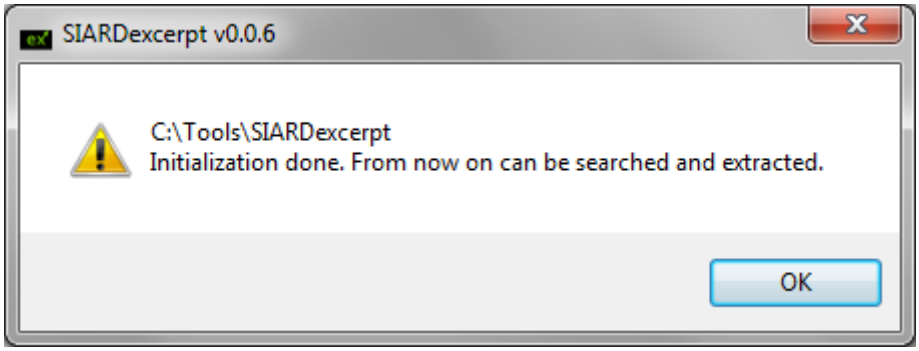
"SIARDexcerptAutoXSL.txt" is used for automatic compilation of a database-specific stylesheet. "taxar.xsl“ allows the display of a single tax register record in the form of a simplified tax declaration.

7 Using SIARDexcerpt

!	SIARDexcerpt is not thread safe! That is to say that concurrent instances of SIARDexcerpt cannot be executed without interfering with each other. Concurrent execution of SIARDexcerpt may lead to errors such as a missing working copy.

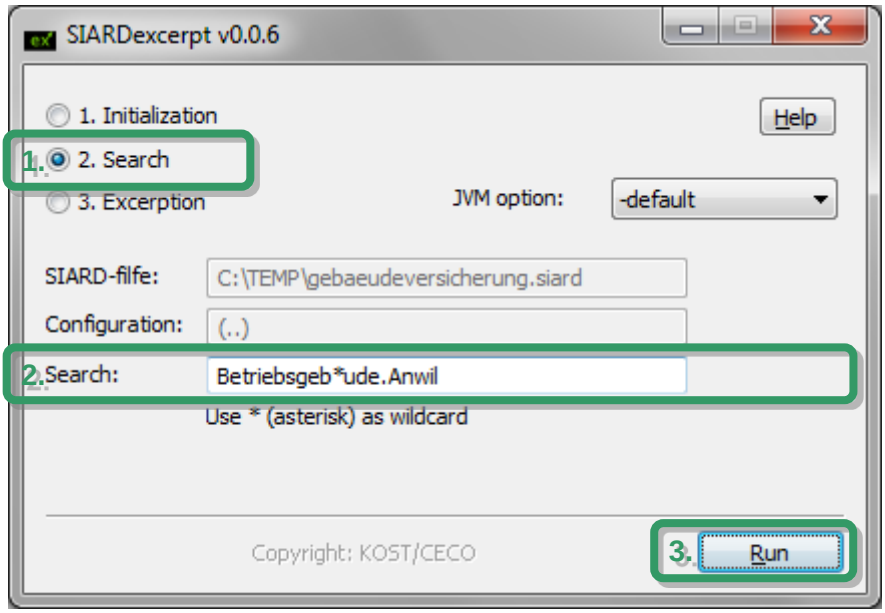
7.1 Search and extraction with "SIARDexcerpt_en.exe" GUI

1	Double click on "SIARDexcerpt_en.exe" in the folder "SIARDexcerpt" to start SIARDexcerpt. 

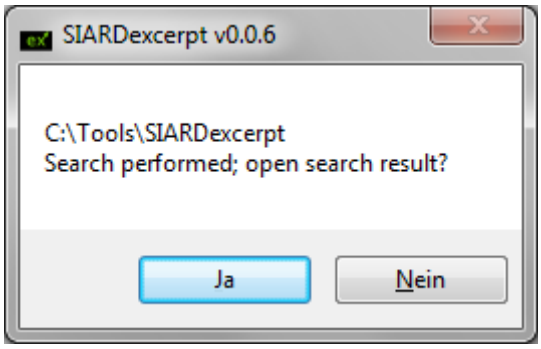
2	Initialisation 1. Initialisation needs to be performed as the first step. 2. Choose or input the path to the SIARD file. 3. Indicate configuration information: • Choose or input the path to the configuration file. • Indicate (..) if no information shall be specified. • Input table name of the chosen main table directly. 4. Start initialisation 
	On finishing initialisation SIARDexcerpt informs that search and extraction can now be started.  Hint: On initialisation a copy of every configuration file is saved as "SIARDexcerpt.conf.xml" in the "configuration" folder, whereby all "(..)" values are filled with values taken from the SIARD file's metadata.xml file. It is recommended to copy this file after initialisation under another name. This enables users to subsequently improve the configuration for the next use of SIARDexcerpt.

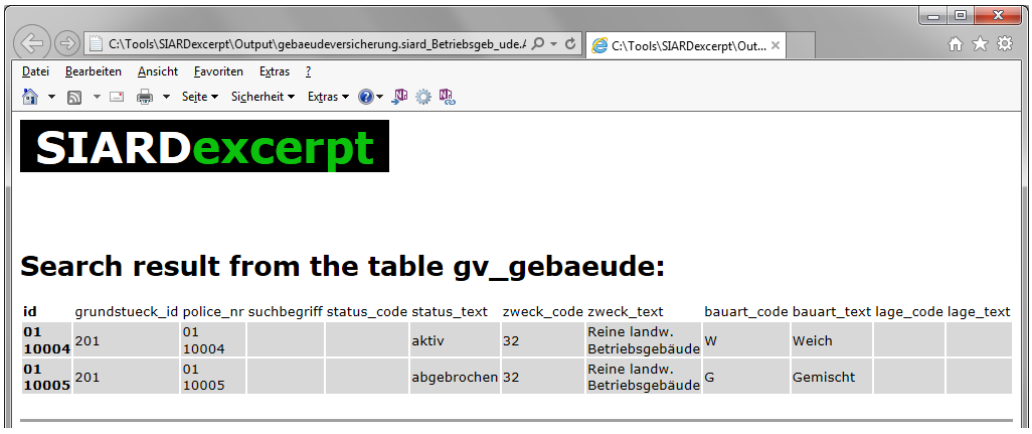
3 Search

1. Select "Search".
2. Input the values to be searched. Individual values must be separated with the wild-card character *. The values' order is significant. Spaces and special characters must be replaced by the wild-card character *.
3. Start search.

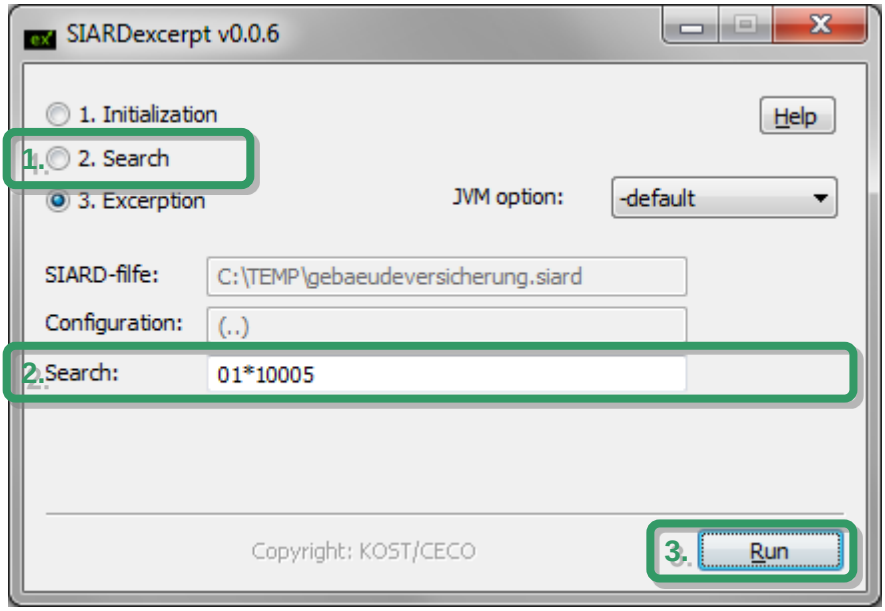
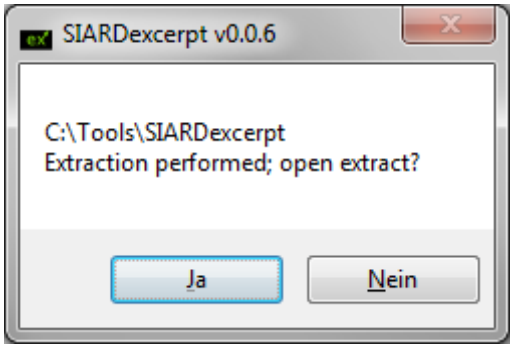


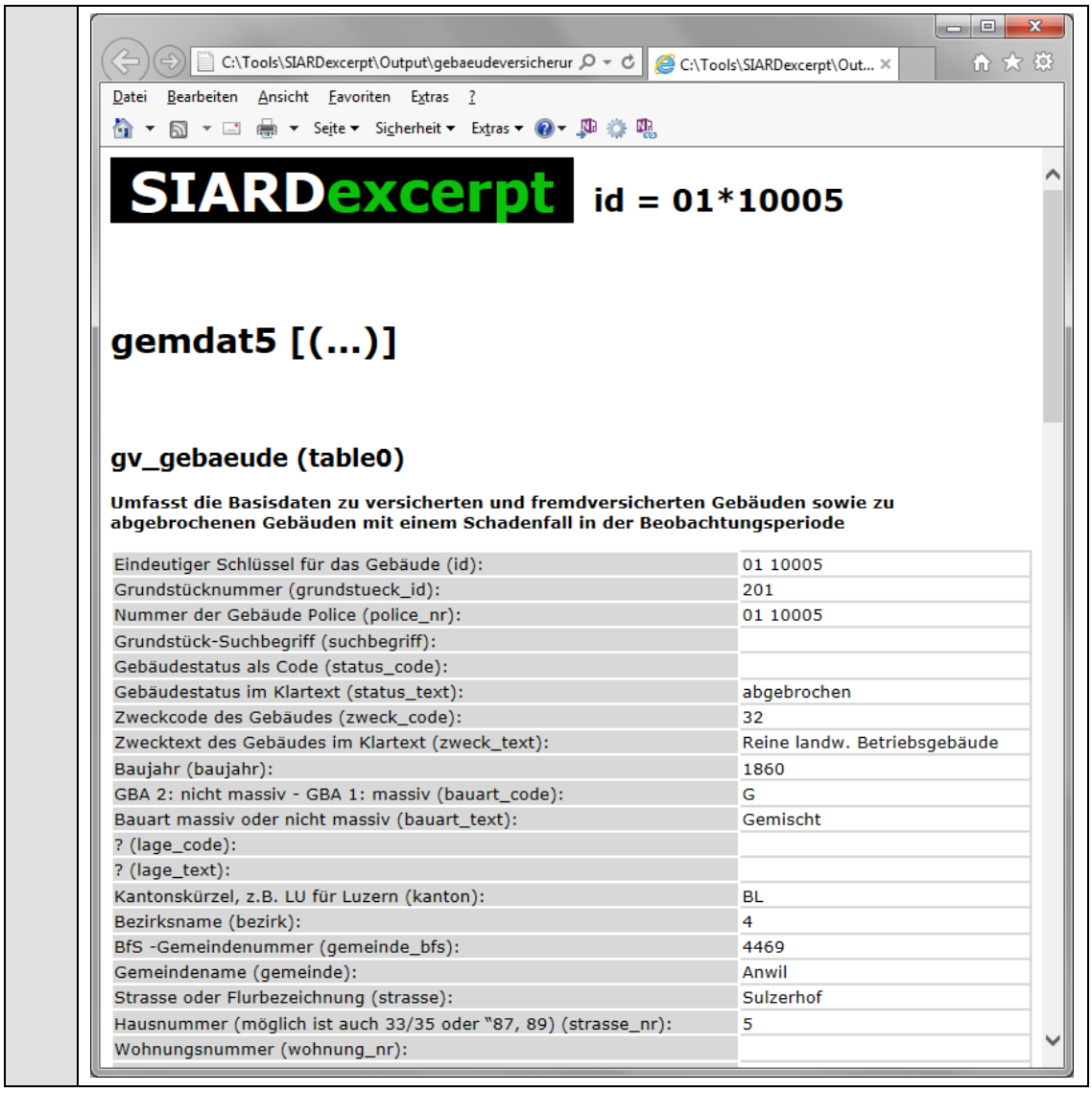
On finishing search, SIARDexcerpt informs that search is completed and proposes to display the search result.





The primary key that is formatted in bold in the result table serves as input for the extraction.

4	Extraction 1. Select "Extraction". 2. Input the primary key. Spaces and other special signs must be replaced by the wild-card character *. 3. Start extraction.  On finishing extraction, SIARDexcerpt informs that extraction is completed and proposes to display the extraction result. 


C:\Tools\SIARDexcerpt\Output\gebaeudeversicherur
C:\Tools\SIARDexcerpt\Out...
Datei Bearbeiten Ansicht Favoriten Extras ?
Seite Sicherheit Extras
SIARDexcerpt
id = 01*10005
gemdat5 [(...)]
gv_gebaeude (table0)
Umfasst die Basisdaten zu versicherten und fremdversicherten Gebäuden sowie zu abgebrochenen Gebäuden mit einem Schadenfall in der Beobachtungsperiode
Eindeutiger Schlüssel für das Gebäude (id): 01 10005
Grundstücknummer (grundstueck_id): 201
Nummer der Gebäude Police (police_nr): 01 10005
Grundstück-Suchbegriff (suchbegriff):
Gebäudestatus als Code (status_code):
Gebäudestatus im Klartext (status_text): abgebrochen
Zweckcode des Gebäudes (zweck_code): 32
Zwecktext des Gebäudes im Klartext (zweck_text): Reine landw. Betriebsgebäude
Baujahr (baujahr): 1860
GBA 2: nicht massiv - GBA 1: massiv (bauart_code): G
Bauart massiv oder nicht massiv (bauart_text): Gemischt
? (lage_code):
? (lage_text):
Kantonskürzel, z.B. LU für Luzern (kanton): BL
Bezirksname (bezirk): 4
BfS -Gemeindenummer (gemeinde_bfs): 4469
Gemeindename (gemeinde): Anwil
Strasse oder Flurbezeichnung (strasse): Sulzerhof
Hausnummer (möglich ist auch 33/35 oder "87, 89) (strasse_nr): 5
Wohnungsnummer (wohnung_nr):

7.2 Manual search and extraction

1

Open a command prompt (Start → Execute… → cmd [RETURN]) and change to the desired working directory (CD C:\Tools\SIARDexcerpt)[4].

General remarks:

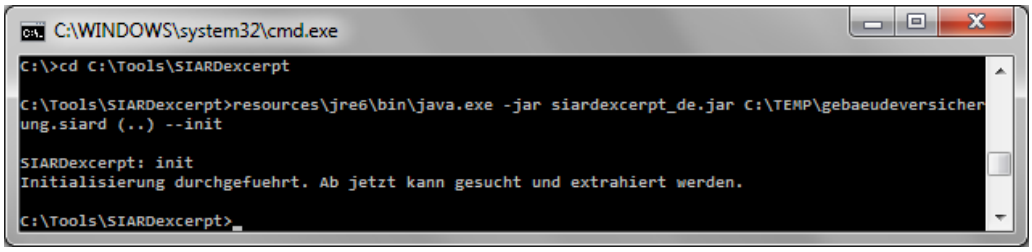
If a file path contains spaces it must be enclosed in quotes.

SIARDexcerpt can be invoked from any location, as long as the program command and the configuration file contain absolute paths.

2

Initialisation always needs to be performed as the first step.

Invoke the SIARDexcerpt command (separate command options with spaces).



A resources\jre6\bin\java.exe[5] -jar
B siardexcerpt_en.jar
C C:\TEMP\gebaeudeversicherung.siard
D (..)
E --init

A = Java command (resources\jre6\bin\java.exe -jar),
B = path and file name siardexcerpt_en.jar,
C = path and file name of the SIARD file to be analysed,
D = configuration file or (..) or name of the primary table,
E = --init.

Initialisation is completed as soon as "Initialisation completed" is displayed. The exit status can take the following values:

0 everything is ok
1 incorrect program call
2 problems at the initialisation stage

Hint:

On initialisation a copy of every configuration file is saved as "SIARDexcerpt.conf.xml" in the "configuration" folder, whereby all "(..)" values are filled with values taken from the SIARD file's metadata.xml file. It is recommended to copy this file after initialisation under another name. This enables users to subsequently improve the configuration for the next use of SIARDexcerpt.

[4] To change the drive type, e.g., CD /D.

[5] Invoking `java -jar` is possible only if Java Runtime Environment (JRE) version 6 is the standard version. JRE7 can cause crashes. SIARDexcerpt is massively slower in connection with JRE8.

3	At search the individual values must be separated with the wild-card character *. The values' order is significant. Spaces and special characters must be replaced by the wild-card character *.  **A** resources\jre6\bin\java.exe -jar **B** siardexcerpt_en.jar **C** C:\TEMP\gebaeudeversicherung.siard **D** (..) **E** --search Betriebsgeb*ude*Anwil **A** = java command (resources\jre6\bin\java.exe -jar), **B** = path and file name siardexcerpt_en.jar, **C** = path and file name of the SIARD file to be analysed, **D** = configuration file or (..) or name of the primary table, **E** = --search search text. Search is completed as soon as "search completed" is displayed. The search results are saved in the indicated output file (gebaeudeversicherung.siard_Betriebsgeb_ude_Anwil_SIARDsearch.xml). The exit status can take the following values: 0 everything is ok, at least one result has been found 1 incorrect program call 2 problems at the initialisation stage

4

At extraction spaces and other special signs in the primary key must be replaced by the wild-card character *.

A resources\jre6\bin\java.exe -jar
B siardexcerpt_en.jar
C C:\TEMP\gebaeudeversicherung.siard
D (..)
E --excerpt 01*10005

A = Java command (resources\jre6\bin\java.exe -jar),
B = path and file name siardexcerpt_en.jar,
C = path and file name of the SIARD file to be analysed
D = configuration file or (..) or name of the primary table
E = --excerpt primary key.

Extraction is completed as soon as "extraction completed" is displayed. The search results are saved in the indicated output file (gebaeudeversicherung.siard_01_10005_SIARDexcerpt.xml). The exit status can take the following values:

0 everything is ok, record extracted successfully
1 incorrect program call
2 problems at the initialisation stage

5

At the end all temporary files are deleted with the "--finish" option.

On initialisation a copy of every configuration file is saved as "SIARDexcerpt.conf.xml" in the "configuration" folder, whereby all "(..)" values are filled with values taken from the SIARD file's metadata.xml file. It is recommended to copy this file after initialisation under another name. This enables users to subsequently improve the configuration for the next use of SIARDexcerpt.

A resources\jre6\bin\java.exe -jar
B siardexcerpt_en.jar
C C:\TEMP\gebaeudeversicherung.siard
D (..)
E --finish

A = Java command (resources\jre6\bin\java.exe -jar),
B = path and file name siardexcerpt_en.jar,
C = path and file name of the SIARD file to be analysed
D = configuration file or (..) or name of the primary table,
E = --finish.

8 Copyright

SIARDexcerpt has been developed by KOST. All rights reserved. SIARDexcerpt has been published by KOST in 2016 under a GNU General Public License v3+.

Notice:	This product includes software developed by the Apache Software Foundation (http://www.apache.org/).

SIARDexcerpt uses the following unmodified components of other manufacturers by embedding them directly into the source code:

Third party application / component	Version	License
Apache Commons http://commons.apache.org/ - commons-collections-3.2.1.jar - commons-configuration-1.6.jar - commons-digester-1.8.jar - commons-lang-2.4.jar - commons-logging-1.1.1.jar	 3.2.1 1.6 1.8 2.4 1.1.1	Apache License 2.0
Apache log4j http://logging.apache.org/log4j/	1.2.12	Apache License 2.0
Apache Xalan-Java http://xml.apache.org/xalan-j/	2.7.0	Apache License 2.0
Apache Xerces http://xerces.apache.org/	2.7.1	Apache License 2.0
Jdom 2.0.0 http://www.jdom.org/	2.0.0	jdom License
Junit 4.4 http://www.junit.org/	4.4	CPL v1.0
Spring Framework API http://static.springsource.org/spring/docs/3.0.x/api/	3.0.0	Apache License 2.0
zip64 http://sourceforge.net/projects/zip64file/	1.02	GPL v2+ License

SIARDexcerpt uses the following unmodified components of other manufacturers which are delivered with SIARDexcerpt:

Third party application / component	Version	License
GREP http://www.gnu.org/software/grep	2.4.2	GPL v3+ License
NSIS v2.46 http://nsis.sourceforge.net/Main_Page	2.46	zlib/libpng License
XML.nsh http://nsis.sourceforge.net/XML_plug-in	2.0	zlib/libpng License
XTrans http://sourceforge.net/projects/xtrans/	1.8.0.2	GPL v2 License

Users of SIARDexcerpt are requested to adhere to these components‘ terms of licence available in the folder SIARDexcerpt\license.

9 Annex

9.1 Description of the configuration file

The configuration file consists of several parts that are described in detail below.

9.1.1 General

<pathtooutput>

Path to output folder of SIARDexcerpt. The initial value is **Output**. It is possible to specify either an absolute or a relative path (to siardexcerpt_en.jar, as in the initial value). The user doesn't need to create this folder.

<pathtoworkdir>

Path to temporary working directory of SIARDexcerpt. The initial value is **temp_SIARDexcerpt**. It is possible to specify either an absolute or a relative path (to siardexcerpt_en.jar, as in the initial value).The user doesn't need to create this folder.

<pathtoxsl>

Path to the extraction stylesheet. The initial value is **(..)**.It is possible to specify either an absolute or a relative path (to siardexcerpt_en.jar).

If (..) is specified the stylesheet is created automatically according to the number of tables and the number of their columns.

<pathtoxslsearch>

Path to the search stylesheet. The initial value is **resources\SIARDexcerptSearch.xsl**. It is possible to specify either an absolute or a relative path (to siardexcerpt_en.jar).

<archive>

Name of the Archives. The initial value is **Archiv**. This is displayed in the search result footer and together with the extracted record.

9.1.2 maintable part

<mainfolder>

Number of the table in which the search is to be executed, i.e. the main table for the extraction. The initial value is **(..)**.

If (..) is specified the table is determined according to the number of primary keys:

1. *Table with the only primary key*
2. *Table with the most used primary key*
3. *If no primary key has been identified, table with the most columns*

<mainname>

Name of the table in which the search is to be executed, i.e. the main table for the extraction. The initial value is **(..)**.

If (..) is specified the table is determined according to the same procedure as <mainfolder>.

<title>

Title of search results. The initial value is **(..)**.

If (..) is specified the string "search result from table <mainname>" is used.

<primarykeyname>

Name of the primary key. The initial value is **(..)**.

If (..) is specified the primary key is determined according to the same procedure as <mainfolder>.

<primarykeycell>

Indication of the cell in which the primary key is stored. The initial value is **(..)**.

If (..)is specified the primary key is determined according to the same procedure as <mainfolder>.

<cellname1> … <cellname11>

Names of the 11 additional columns for the output of the search result. The initial values are **(..)**.

If (..) is specified the columns are prioritised according to their data type as follows. The first 11 or 12 columns are selected.

1. *Top priority (1) for "CHARACTER VARYING", "CHARACTER" and "DATE"*
2. *Priority 2 for "DECIMAL", "NATIONAL CHARACTER VARYING" and "NATIONAL CHARACTER"*
3. *Priority 3 for "BIGINT", "INTEGER", "SMALLINT" and "NUMERIC"*
4. *Priority 4 for "DOUBLE PRECISION", "FLOAT", "INTERVAL" and "REAL"*
5. *Priority 5 for "TIME", "TIME WITH TIME ZONE", "TIMESTAMT" and "TIMESTAMP WITH TIME ZONE"*
6. *Priority 6 for "BINARY VARYING", "BINARY", "BIT VARYING", "BIT" and "XML"*
7. *Priority 7 for "BINARY LARGE OBJECT", "BOOLEAN", "CHARACTER LARGE OBJECT"*
8. *If necessary "NATIONAL CHARACTER LARGE OBJECT" and other types can be used.*

<cellnumber1> … < cellnumber11>

Indication of the cells in which the columns for the output of the search result are stored. The initial values are **(..)**.

If (..) is specified the cell numbers are determined according to the same procedure as <cellname>.

9.1.3 subtables part

<subtables>

Information[6] about the tables to be joined and the respective keys. The initial value is **(..)**.

If (..) is selected, <name>, <folder>, and <foreignkeycell> are determined for every table <subtable> that references the primary key as foreign key.

[6] For every <subtable>, <name>, <folder>, and <foreignkeycell> are necessary and will be identified.

9.2 Functional principle of SIARDexcerpt