



## Manual

### Content

1	Preface .....	2
2	System requirements .....	2
3	Open issues / Feedback .....	2
4	Installation .....	3
5	Configuration of SIARDexcerpt, of SIARD search, and extraction .....	3
5.1	Parts of the configuration file .....	4
6	Resources of SIARDexcerpt .....	5
7	Using SIARDexcerpt .....	6
7.1	Search and extraction using "SIARDexcerpt.exe" GUI .....	6
7.2	Manual search and extraction .....	10
8	Copyright .....	12
9	Annex .....	13
9.1	Description of the configuration file .....	13
9.2	Functional principle of SIARDexcerpt .....	15

## 1 Preface

SIARDexcerpt is a java-based application for searching and extracting individual records from SIARD files<sup>1</sup>. The single record is converted into a human readable form using a custom or generic stylesheet. This application is licensed under the GPL3+ license and is made open source available to the public by KOST. SIARDexcerpt relies on unmodified components from other vendors, which are directly included in the source code of SIARDexcerpt. Users of SIARDexcerpt are required to follow the license terms of all such components. Detailed information can be found in chapter 8.

SIARDexcerpt fulfills the requirements described below, which were presented in chapter 9.2 as a functional principle.

Initialization: During initialization, the SIARD file is unpacked in the work directory. In addition the desired configuration is copied to the predefined place. If required, the configuration is automatically filled in as described in chapter 9.1 using metadata.xml and temporarily saved as SIARDexcerpt.conf.xml.

Search: After initialization the matching lines are searched with grep. The \* (asterisk) serves as a wildcard for the search. SIARDexcerpt copies the lines and 12 predefined columns are output as a preview. Through a stylesheet the table can be displayed and read directly in the explorer. The search result is stored in the output folder.

Extraction: As soon as the master key is known, the extraction can be started directly. The extract is stored in the output folder. A stylesheet can be used to display and read the data set directly in the Explorer.

Exit: Finally, the temporary configuration SIARDexcerpt.conf.xml and the unpacked SIARD file are deleted again.

## 2 System requirements

- Microsoft Windows 98 and newer
- At least 128 MB RAM (512 MB or more recommended)
- At least 20 GB hard disk space
- Java Runtime Environment (JRE) version 8 respectively 1.8 incl. JavaFX or Liberica Full openJDK<sup>2</sup>

## 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](mailto:kost-val@kost-ceco.ch). These issues are managed by the development team. Any and all contributions are welcome.

---

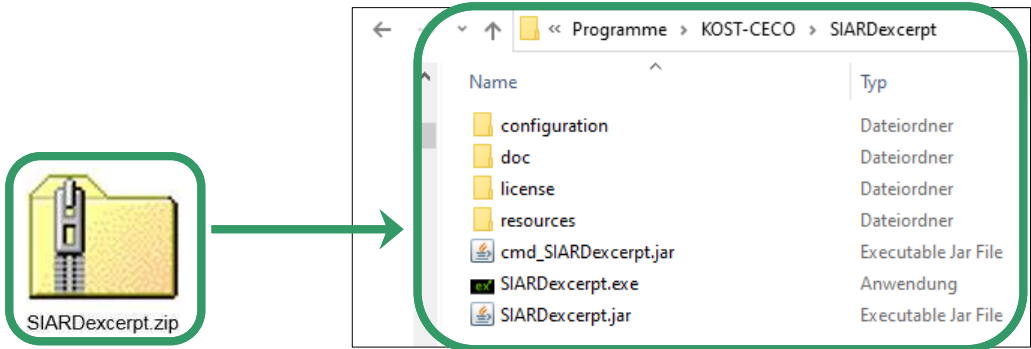
<sup>1</sup> The SIARD specification (*Software Independent Archiving of Relational Databases*) can be downloaded from the eCH website:

<http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0165>.

<sup>2</sup> The following version was successfully tested: <https://download.bell-sw.com/java/14.0.2+13/bellsoft-jdk14.0.2+13-windows-amd64-full.msi>

## 4 Installation

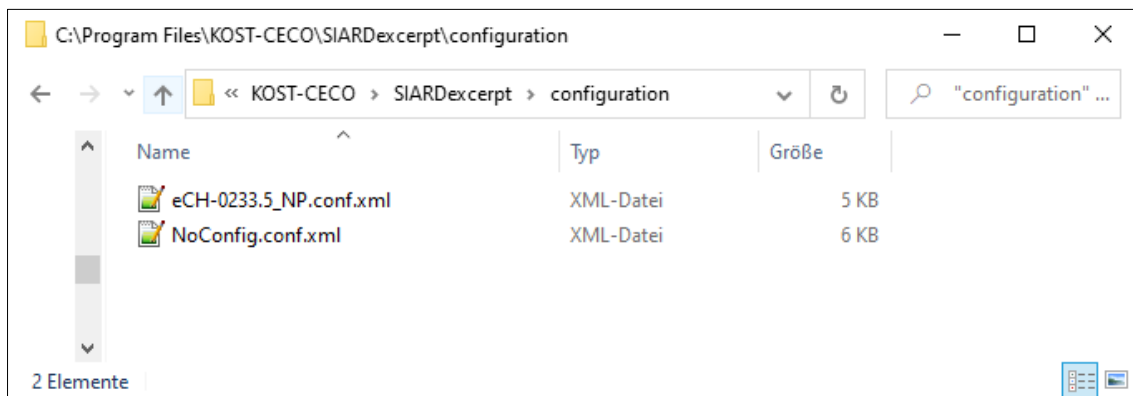
**1** Download SIARDexcerpt and unzip it to the desired directory.



Name	Typ
configuration	Dateiordner
doc	Dateiordner
license	Dateiordner
resources	Dateiordner
cmd_SIARDexcerpt.jar	Executable Jar File
SIARDexcerpt.exe	Anwendung
SIARDexcerpt.jar	Executable Jar File

## 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 ++<sup>3</sup> should be used to modify XML files.



"NoConfig.conf.xml" in the "configuration" folder serves as the basis. This can be used for all SIARD files. The configuration file "eCH-0233.5\_NP.conf.xml" is included as an example of a special configuration. The latter is used for extraction from the assessment database or the assessment statement.

During initialization a copy of each configuration is created as "SIARDexcerpt.conf.xml" in the directory "USERHOME\ siardexcerpt\configuration" and all "(..)" are filled with values derived from metadata.xml of the SIARD file. It is recommended to copy this after initialization and save it under a different name. Corrections in the configuration can then be made and used directly the next time.

<sup>3</sup> Portable Notepad ++ is available from [http://portableapps.com/de/apps/development/notepadpp\\_portable](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 delivered configuration allows instant search and extraction from any SIARD files. The components are briefly described below.

### 5.1.1 General

xml-Tag	Description: default value
<pathtoxsl>	Path to the extraction stylesheet, e.g.: <b>(..)</b>
<pathtoxslsearch>	Path to the search stylesheet: <b>resources\SIARDexcerptSearch.xsl</b>
<archive>	Name of the Archives: <b>Archiv</b>
<insensitive>	Ignore upper and lower case: <b>yes</b>
<sed>	Specify whether sed should be used to put all the individual data of the row on one line: <b>no</b>

### 5.1.2 maintable part

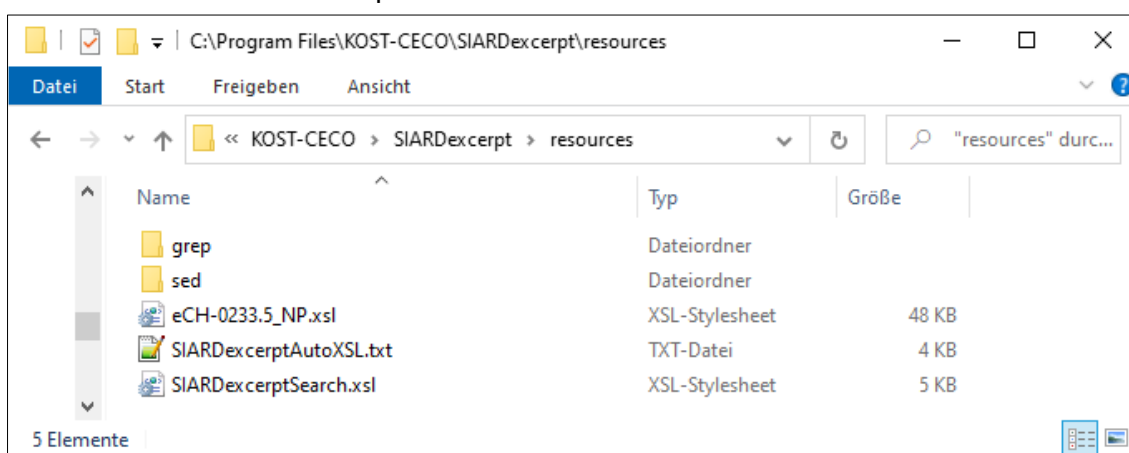
xml-Tag	Description: default value
<mfolder>	Number of the table in which the search is to be executed, i.e. the main table for the extraction: <b>(..)</b>
<mname>	Name of the table in which the search is to be executed, i.e. the main table for the extraction: <b>(..)</b>
<mtitle>	Title of search results: <b>(..)</b>
<mpkname>	Name of the primary key: <b>(..)</b>
<mpkcell>	Indication of the cell in which the primary key is stored: <b>(..)</b>
<mschemaname>	Schema name in which the main table is located: <b>(..)</b>
<mschemafolder>	Schema folder in which the main table is located: <b>(..)</b>
<mc1name> ... <mc11name>	Names of the additional columns for the output of the search result: <b>(..)</b>
<mc1number> ... <mc11number>	Indication of the cells in which the columns for the output of the search result are stored: <b>(..)</b>

### 5.1.3 subtables part

xml-Tag	Description: <b>default value</b>
<st1keyname> <st1name> <st1folder> <st1fkcell> <st1schemafolder> <st1schemaname>	Information <sup>4</sup> about the tables to be joined and the respective keys: (..) <b>Up to 20 tables can be linked.</b>

## 6 Resources of SIARDexcerpt

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



"grep" and "sed" are required by SIARDexcerpt.

Through the stylesheet "SIARDexcerptSearch.xsl" the table with the search results can be displayed and read directly in the explorer.

"SIARDexcerptAutoXSL.txt" is used for the automatic creation of a database-specific stylesheet. "eCH-0233.5\_NP.xsl" allows the single assessment statement record to be displayed in a simplified form of a tax return.

---

<sup>4</sup> For every <subtable>, <name>, <folder>, <foreignkeycell>, <schemafolder> and <schemaname> are necessary and will be identified.

## 7 Using SIARDexcerpt

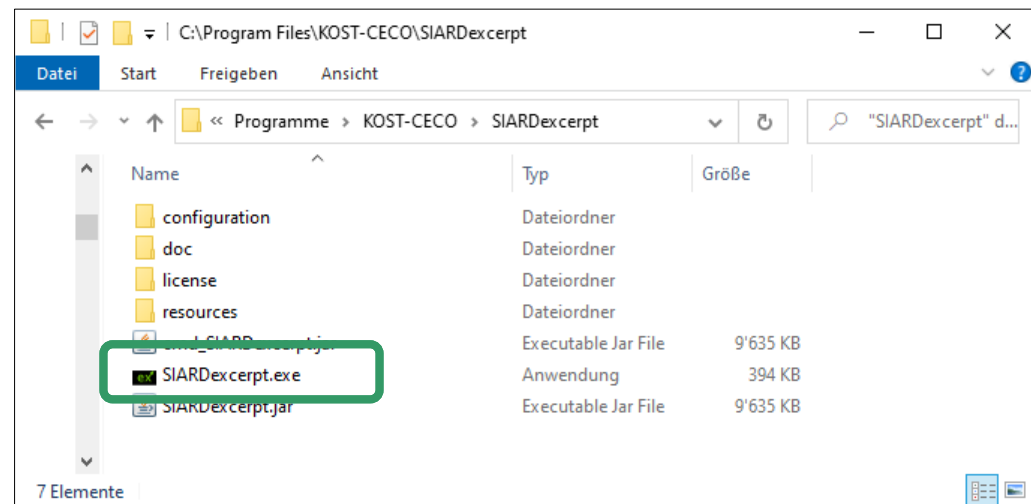


SIARDexcerpt is not thread-safe!

This means that multiple instances of SIARDexcerpt cannot be run simultaneously without interfering with each other. If SIARDexcerpt is executed at the same time, errors such as a missing working copy may occur.

### 7.1 Search and extraction using "SIARDexcerpt.exe" GUI

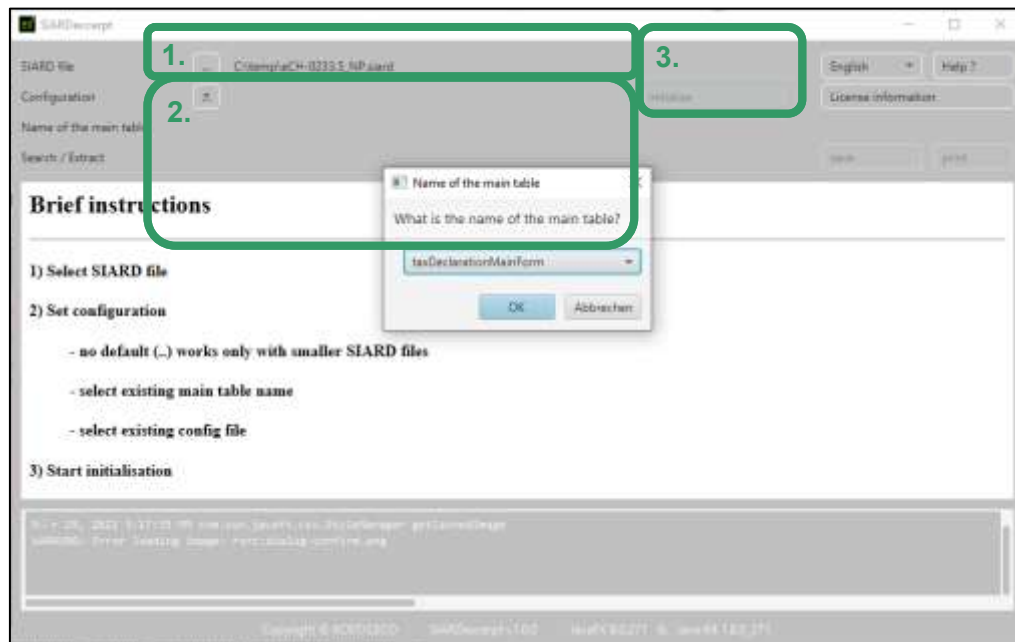
- 1 Start SIARDexcerpt by double-clicking on "SIARDexcerpt.exe"<sup>5</sup> in the "SIARDexcerpt" folder.



<sup>5</sup> If openJDK<sup>2</sup> is used, the GUI must be started via "SIARDexcerpt.jar" and "Open with...". The java.exe of openJDK must be specified.

**2** Initialization: The first thing that must always be done is the initialization.

1. Select the path to SIARD file or enter it directly.
2. Enter information about the configuration:
  - Select the path to the configuration file
  - (..)if no specifications are made
  - Enter table name of the desired main table directly
3. Start initialization.



At the end of the initialization it is indicated that now searching and extracting is possible.



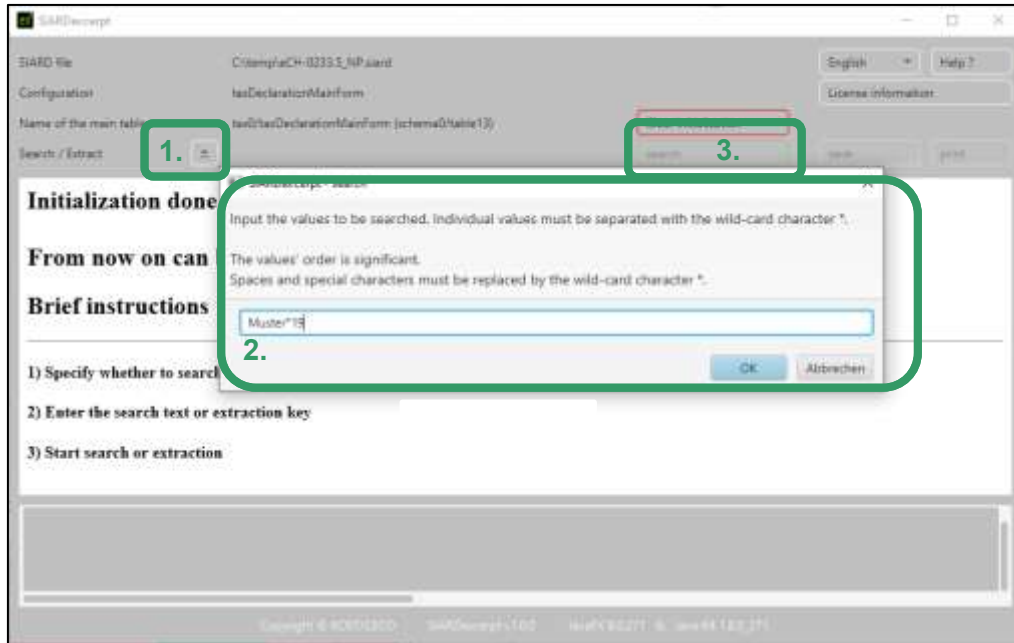
**Notice:**

During initialization, a copy of each configuration is created as "SIARDexcerpt.conf.xml" in the directory "USERHOME/. siardexcerpt/configuration" and all "(..)" are created with values derived from metadata.xml of the SIARD file. It is recommended to copy these after initialization and save them under a different name. Corrections in the configuration can then be made and used directly the next time.

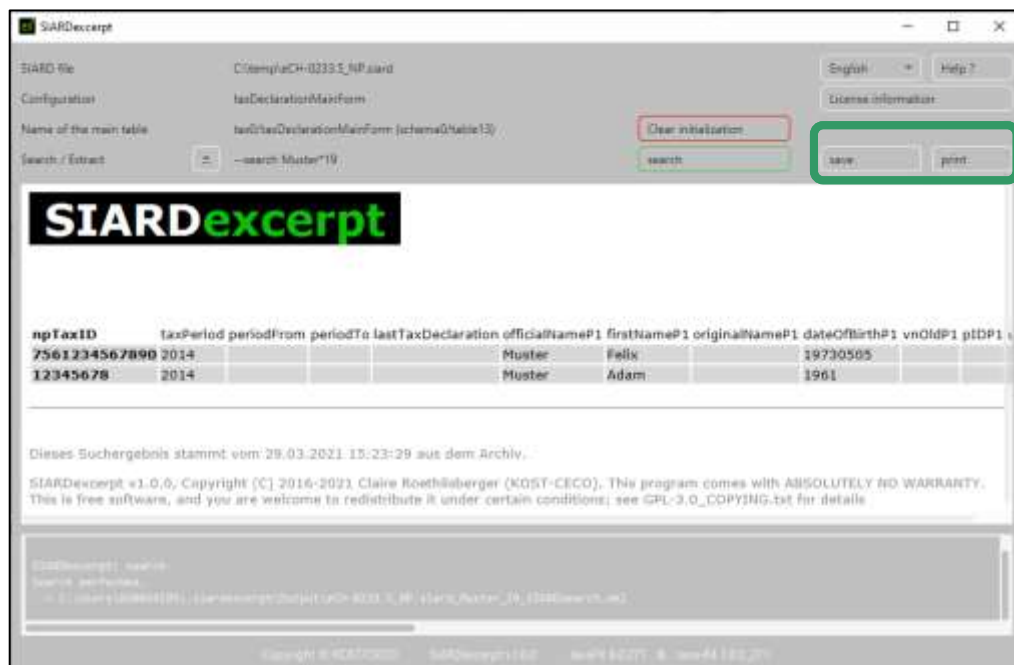
### 3

#### Search

1. Select search.
2. Enter the values you are looking for. The individual values must be separated with the wildcard \*. The sequence is taken over. Spaces and other special characters must be replaced by the wildcard \*.
3. Start search



At the end of the search, it is indicated that the search has been performed and the result is displayed. If desired, the result can be saved elsewhere or printed.

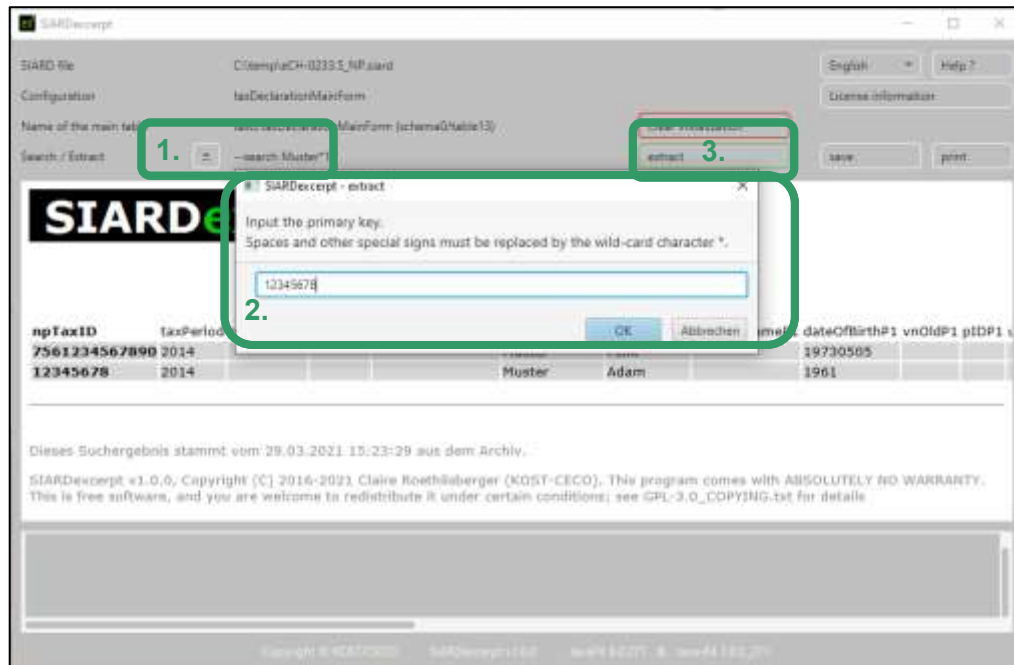


Based on the primary key, which is displayed in bold in the table, can be extracted.



#### 4 Extraction

1. Select extraction.
2. Enter the key. Spaces and other special characters must be replaced by the wildcard \*.
3. Start extraction

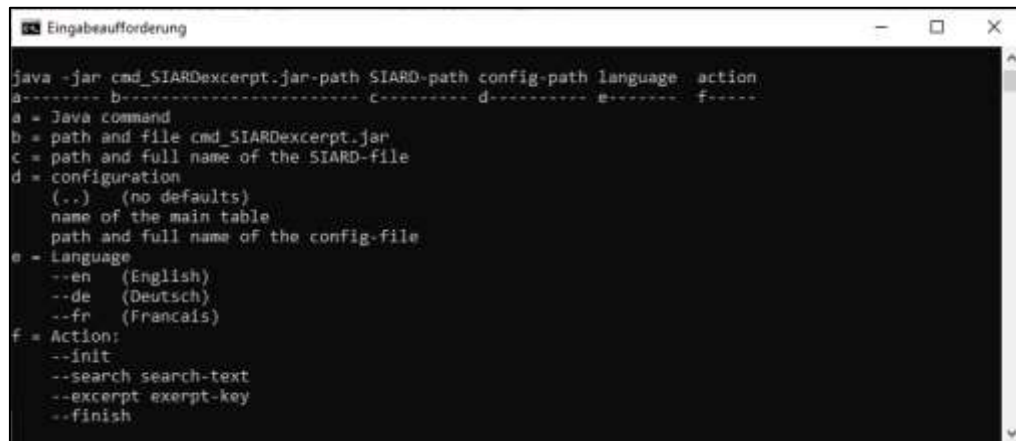


At the end of the extraction, it is indicated that the extraction has been performed and the result is displayed. If desired, the result can be saved elsewhere or printed.



## 7.2 Manual search and extraction

- 1 Open command prompt (Start Run → cmd → [OK]) and change to the desired working directory (CD "C:\Program Files\KOST-CECO\SIARDexcerpt"). Structure SIARDexcerpt command:

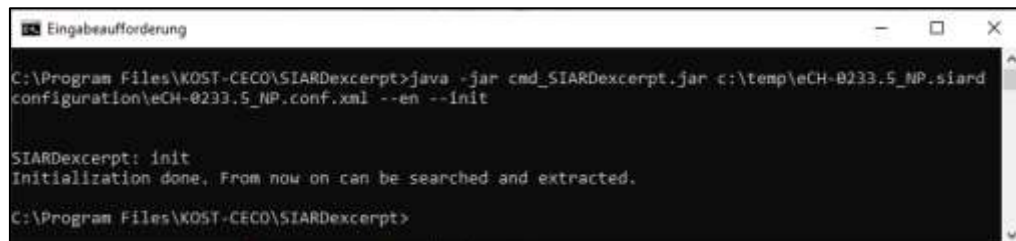


```
java -jar cmd_SIARDexcerpt.jar-path SIARD-path config-path language action
a----- b----- c----- d----- e----- f-----
a = Java command
b = path and file cmd_SIARDexcerpt.jar
c = path and full name of the SIARD-file
d = configuration
  (..) (no defaults)
  name of the main table
  path and full name of the config-file
e = Language
  --en (English)
  --de (Deutsch)
  --fr (Francais)
f = Action:
  --init
  --search search-text
  --excerpt excerpt-key
  --finish
```

### General notes:

If a path contains spaces, it must be entered in leading and trailing characters. SIARDexcerpt can also be called from any location. However, this requires the input of absolute paths not only in the program call itself, but also in the configuration file.

- 2 The first thing that must always be done is the initialization. Start SIARDexcerpt program call (separate the individual input components with spaces).



```
C:\Program Files\KOST-CECO\SIARDexcerpt>java -jar cmd_SIARDexcerpt.jar c:\temp\ech-0233.5_NP.siard
configuration\ech-0233.5_NP.conf.xml --en --init

SIARDexcerpt: init
Initialization done, From now on can be searched and extracted.

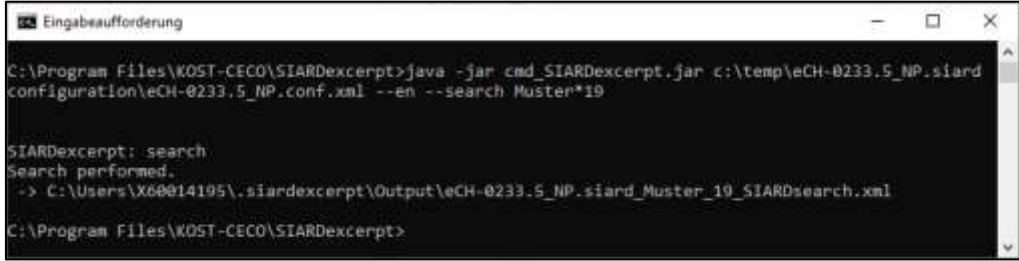


C:\Program Files\KOST-CECO\SIARDexcerpt>
```

Initialization has been performed as soon as the text "Initialization done" appears. The *exit* status can assume the following values:

- 0 everything OK
- 1 Error in program call
- 2 problems have occurred during initialization

### Notice:

During initialization, a copy of each configuration is created as "SIARDexcerpt.conf.xml" in the directory "USERHOME/. siardexcerpt/configuration" and all "(..)" are created with values derived from metadata.xml of the SIARD file. It is recommended to copy these after initialization and save them under a different name. Corrections in the configuration can then be made and used directly the next time.

3	<p>When searching, the individual values must be separated with the wildcard *. The order is taken over. Spaces and other special characters must be replaced by the wildcard *.</p>  <p>The search is finished as soon as the text "Search completed" appears. The search results can be seen in the output file (eCH-0233.5_NP.siard_Muster_19_SIARDsearch.xml). The exit status can take the following values:</p> <ul style="list-style-type: none"> <li>0 everything OK, at least one result was found</li> <li>1 Error in program call</li> <li>2 problems have occurred during the search</li> </ul>
4	<p>During extraction, spaces and other special characters in the key must be replaced by the wildcard * if present.</p>  <p>The extraction is finished as soon as the text "Extraction completed" appears. The extracted record is shown in the output file (eCH-0233.5_NP.siard_12345678_SIARDexcerpt.xml) ersichtlich. The <i>exit status</i> can take the following values:</p> <ul style="list-style-type: none"> <li>0 everything OK, record could be extracted</li> <li>1 Error in program call</li> <li>2 problems have occurred during extraction</li> </ul>
5	<p>Finally, "--finish" will delete all temporary files.</p> <p>During initialization, a copy of each configuration is created as "SIARDexcerpt.conf.xml" in the directory "USERHOME/. siardexcerpt/configuration" and all "(.)" are created with values derived from metadata.xml of the SIARD file. It is recommended to copy these after initialization and save them under a different name. Corrections in the configuration can then be made and used directly the next time.</p> 

## 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+.

<b>Notice:</b>	This product includes software developed by the Apache Software Foundation ( <a href="http://www.apache.org/">http://www.apache.org/</a> ).
----------------	---

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

Third party application / component	Version	License
<b>Apache Commons</b> <a href="http://commons.apache.org/">http://commons.apache.org/</a>		Apache License 2.0
- commons-io-2.6.jar	2.6	
- commons-logging-1.2.jar	1.2	
<b>Apache log4j</b> <a href="http://logging.apache.org/log4j/">http://logging.apache.org/log4j/</a>	1.2.12	Apache License 2.0
<b>Apache Xerces</b> <a href="http://xerces.apache.org/">http://xerces.apache.org/</a>	2.7.1	Apache License 2.0
<b>Junit 4.12</b> <a href="http://www.junit.org/">http://www.junit.org/</a>	4.12	CPL v1.0
<b>Spring Framework API</b> <a href="http://static.springframework.org/spring/docs/3.0.x/api/">http://static.springframework.org/spring/docs/3.0.x/api/</a>	5.0.8	Apache License 2.0
<b>zip64</b> <a href="http://sourceforge.net/projects/zip64file/">http://sourceforge.net/projects/zip64file/</a>	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
<b>GNU grep</b> <a href="http://www.gnu.org/software/grep">http://www.gnu.org/software/grep</a>	2.4.2	GPL v3+ License
<b>GNU sed</b> <a href="https://www.gnu.org/software/sed">https://www.gnu.org/software/sed</a>	4.4	GPL v3+ 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 part

##### <pathtoxsl>

Path specification to the stylesheet of the extraction. The initial value is **(..)**. The absolute path can be specified or the relative path (from the point of view of cmd\_SIARDexcerpt.jar).

*With (..) the XSL is created automatically based on the number of tables and their number of columns.*

##### <pathtoxslsearch>

Path specification to the stylesheet of the search. The initial value is **re-sources\SIARDexcerptSearch.xsl**. The absolute path can be specified or, as in the initial value, the relative path (from the point of view of siardexcerpt\_en.jar).

##### <archive>

Name of the archive. The initial value is **Archive**. This is displayed in the footer at the search result and at the extracted record.

##### <insensitive>

Ignore upper and lower case in the search. The initial value is **yes**.

##### <sed>

Specifies whether sed should be used to put all the individual data of the row on one line. The initial value is **no**. Most SIARD programs already write all data of the row on one line.

#### 9.1.2 maintable part

##### <mfolder>

Number of the table to be searched, respectively of the main table during extraction. The initial value is **(..)**.

*For (..) the table is determined according to the number of primary keys:*

- 1. Table with the single primary key*
- 2. Table with the primary key, which is used most often*
- 3. If no primary key the table with the most columns is used*

##### <mname>

Name of the table to be searched in, respectively the main table during extraction. The initial value is **(..)**.

*With (..) the table is determined as with < mfolder>.*

##### <mtitle>

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

*At (..) the text "Search result from table <mname>:" is used.*

### **<mpkname>**

Specification of how the master key is to be designated. The initial value is (..).

### **<mpkcell>**

Specification in which cell the master key is located. The initial value is (..).

### **<mschemaname>**

Schema name in which the main table is located. The initial value is (..).

### **<mschemafolder>**

Schema folder in which the main table is located. The initial value is (..).

### **<mc1name> ... <mc11name>**

Specification of how the eleven further columns for the output of the search result are to be designated. The initial value in each case is (..).

*With (..) the columns are prioritized according to their data type as follows and the first 11 respectively 12 columns are selected:*

1. *Highest 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 if necessary others are used*

### **<mc1number> ... <mc11number>**

Specifies in which cell the columns for the output of the search result are located. The initial value in each case is (..).

## **9.1.3 subtables part**

### **<subtables>**

Information<sup>6</sup> about the tables that are to be linked and the keys. Up to 20 tables can be linked. The initial value is (..).

At (..) the data is determined for each table which references the primarykey as foreignkey.

---

<sup>6</sup> <st1keyname> , <st1name> , <st1folder> , <st1fkcell> , <st1schemafolder> , <st1schema-name>

## 9.2 Functional principle of SIARDxcerpt

