Ein Gemeinschaftsunternehmen von Schweizer Archiven



Anwendungshandbuch

Inhalt

1	Vorwort	2
	Systemvoraussetzungen	
	Offene Punkte / Feedback	
4		
	4.1 Installation von SIARDexcerpt	
5	Konfiguration des SIARDexcerpt und der SIARD-Suche und Extraktion	
	5.1 Bestandteile der Konfigurationsdatei	
6	Ressourcen des SIARDexcerpt	5
7	SIARDexcerpt verwenden	5
	7.1 Suche und Extraktion mittels "SIARDexcerpt_de.exe"-GUI	5
	7.2 Manuelle Suche und Extraktion	10
8	Urheberrecht	13
9	Anhang	14
	9.1 Beschreibung der Konfigurationsdatei	
	9.2 Funktionsprinzin SIARDexcerpt	17

1 Vorwort

SIARDexcerpt ist eine java-basierte Anwendung zur Suche und Extraktion einzelner Records aus SIARD-Dateien (*Software Independent Archiving of Relational Databases*). Der Einzelne Record wird anhand eines benutzerspezifischen oder generischen Stylesheets in eine Menschenlesbare Form gebraucht. Diese Anwendung steht unter der GPL3+ Lizenz und wird durch die KOST der Öffentlichkeit quelloffen zur Verfügung gestellt. SIARDexcerpt stützt sich auf unveränderte Komponenten anderer Hersteller, welche direkt im Quellcode von SIARDexcerpt eingebunden sind. Die Benutzer von SIARDexcerpt sind gehalten, die Lizenzbestimmungen all dieser Komponenten zu befolgen. Ausführliche Informationen sind im Kapitel 8 ersichtlich.

SIARDexcerpt erfüllt die im Folgenden beschriebenen Anforderungen, welche im Kapitel 9.2 als Funktionsprinzip dargestellt wurde.

<u>Initialisierung:</u> Bei der Initialisierung wird die SIARD-Datei im Workverzeichnis entpackt. Zudem wird die gewünschte Konfiguration an die vordefinierte Stelle kopiert. Bei Bedarf wird die Konfiguration nach bestem Wissen und Gewissen anhand metadata.xml automatisch ausgefüllt und als SIARDexcerpt.conf.xml temporär gespeichert.

<u>Suche:</u> Nach der Initialisierung wird mit grep die passenden Zeilen gesucht. Als Wildcard bei der Suche dient der * (Stern). SIARDexcerpt kopiert die Zeilen und die 12 vordefinierten Spalten werden ausgegeben. Durch ein Stylesheet kann direkt im Explorer die Tabelle angezeigt und gelesen werden. Das Suchresultat wird im konfigurierten Output-Ordner abgespeichert.

<u>Extraktion:</u> Sobald der Hauptschlüssel bekannt ist kann direkt die Extraktion gestartet werden. Das Extrakt wird im konfigurierten Output-Ordner abgespeichert. Durch ein Stylesheet kann direkt im Explorer der Datensatz angezeigt und gelesen werden.

<u>Beenden:</u> Zum Schluss wird die temporäre Konfiguration SIARDexcerpt.conf.xml sowie die entpackte SIARD-Datei wieder gelöscht.

Die Resultate (inklusive Meldungen zu Inkonsistenzen oder Fehler) werden pro Schritt ausgegeben und in eine Logdatei geschrieben.

2 Systemvoraussetzungen

Microsoft Windows 98 und neuer
Mindestens 128 MB RAM (512 MB oder mehr empfohlen)
Mindestens 20 GB Festplattenspeicher
Java Runtime Environment (JRE) Version 6 respektive 1.6

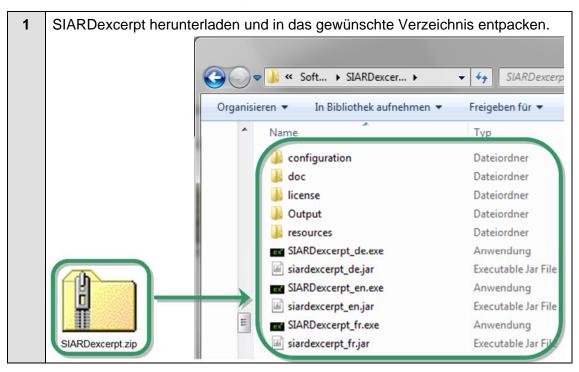
3 Offene Punkte / Feedback

Die offenen Punkte, von Bugs respektive Fehler bis hin zu Ergänzungswünschen und Fragen sind auf der Entwicklungsplattform GitHub unter Issues ersichtlich (https://github.com/KOST-CECO/SIARDexcerpt/issues) und können an kost-val@kost-ceco.ch gemeldet werden.

Diese Liste kann und soll durch jedermann erweitert werden und wird durch das Entwicklerteam bearbeitet.

4 Installation

4.1 Installation von SIARDexcerpt



5 Konfiguration des SIARDexcerpt und der SIARD-Suche und Extraktion

Einige Konfigurationen des SIARDexcerpt müssen vorgängig in einer Konfigurationsdatei definiert werden. Für Modifikationen an XML-Dateien sollte ein XML-Editor oder Notepad ++¹ verwendet werden.



Als Basis dient "NoConfig.conf.xml" im Ordner "configuration". Diese kann für alle SIARD-Dateien verwendet werden. Als Beispiel für eine spezielle Konfiguration ist die Konfigurationsdatei "taxar.conf.xml" mitenthalten. Letztere dient für die Extraktion aus der Steuerregisterdatenbank respektive dem Steuerregisterauszug.

Bei der Initialisierung wird von jeder Konfiguration eine Kopie als "SIARDexcerpt.conf.xml" im Ordner "configuration" erstellt und alle "(..)" mit logischen Werten aus metadata.xml der SIARD-Datei gefüllt. Es empfiehlt sich diese nach der Initialisierung zu Kopieren und unter einem anderen Namen abzuspeichern. Nachbesserungen in der Konfiguration können dann vorgenommen und beim nächsten Mal direkt verwendet werden.

-

¹ Die Portable Notepad ++ Applikation kann unter http://portableapps.com/de/apps/development/notepadpp_portable heruntergeladen werden.

5.1 Bestandteile der Konfigurationsdatei

Die Konfigurationsdatei ist in verschiedenen Teilen aufgebaut, welche im Kapitel 9.1 detailliert beschrieben werden.

Die ausgelieferte Konfiguration ermöglicht eine sofortige Suche und Extraktion aus beliebigen SIARD-Dateien.

Nachfolgend werden die Bestandteile kurz beschrieben.

5.1.1 Allgemeiner Teil

xml-Tag	Beschreibung: Initialwert	
<pathtooutput></pathtooutput>	Pfadangabe zum Outputverzeichnis von SIARDexcerpt: Output	
<pathtoworkdir></pathtoworkdir>	Pfadangabe zum temporären Arbeitsverzeichnis von SIARDexcerpt: temp_SIARDexcerpt	
<pathtoxsl></pathtoxsl>	Pfadangabe zum Stylesheet der Extraktion. In unrem Beispiel ist dies: ()	
<pathtoxslsearch></pathtoxslsearch>	Pfadangabe zum Stylesheet der Suche: resources\SIARDexcerptSearch.xsl	
<archive></archive>	Name des Archivs zum Beispiel: Staatsarchiv Zuerich	

5.1.2 maintable Teil

xml-Tag	Beschreibung: Beispiel	
<mainfolder></mainfolder>	Nummer in welcher Tabelle gesucht werden soll respektive die Haupttabelle bei der Extraktion: ()	
<mainname></mainname>	Name der Tabelle in welcher gesucht werden soll respektive die Haupttabelle bei der Extraktion: ()	
<title></td><td>Titel der Suchergebnisse: ()</td></tr><tr><td><pre><pre><pre><pre><pre><pre><pre><pre></td><td>Angabe wie der Hauptschlüssel bezeichnet werden soll: ()</td></tr><tr><td><pre><pre><pre><pre><pre><pre><pre><pre></td><td>Angabe in welcher Zelle sich der Hauptschlüssel befindet: ()</td></tr><tr><td><cellname1> <cellname11></td><td>Angabe wie die weiteren Spalten für die Ausgabe des Suchergebnisses bezeichnet werden soll: ()</td></tr><tr><td><cellnumber1> < cellnumber11></td><td>Angabe in welcher Zelle sich die Spalten für die Ausgabe des Suchergebnisses befindet: ()</td></tr></tbody></table></title>		

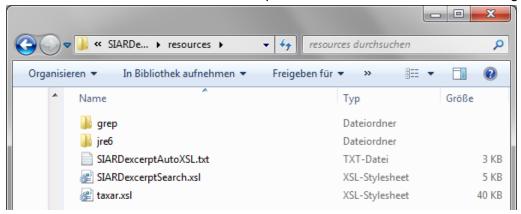
5.1.3 subtables Teil

xml-Tag	Beschreibung: Initialwert	
<subtables></subtables>	Angaben ² über die Tabellen welche verknüpft werden sollen und den Schlüsseln: ()	

² pro <subtable> wird <name>, <folder> und <foreignkeycell> benötigt respektive ermittelt

6 Ressourcen des SIARDexcerpt

Sämtliche Ressourcen des SIARDexcerpt sind im Unterordner "resources" abgelegt.



"grep" und "jre6" werden durch SIARDexcerpt benötigt.

Durch das Stylesheet "SIARDexcerptSearch.xsl" kann direkt im Explorer die Tabelle mit den Suchresultaten angezeigt und gelesen werden.

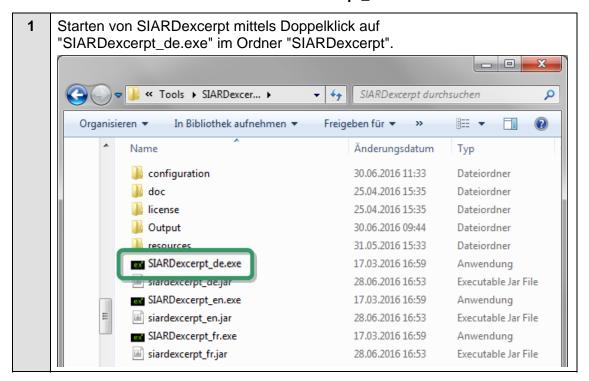
"SIARDexcerptAutoXSL.txt" wird für die automatische Erstellung eines datenbankspezifischen Stylesheets verwendet. "taxar.xsl" ermöglicht die Darstellung des einzelnen Steuerregister-Record in Form einer vereinfachten Steuererklärung.

7 SIARDexcerpt verwenden

SIARDexcerpt ist nicht Thread-sicher!

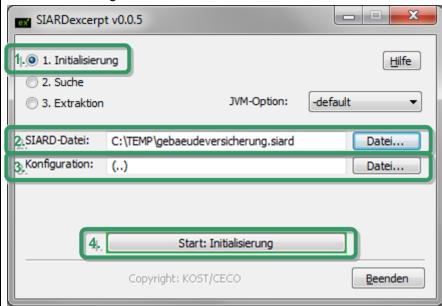
Das bedeutet, dass nicht mehrere Instanzen von SIARDexcerpt gleichzeitig ausgeführt werden können, ohne sich gegenseitig zu behindern. Wird SIARDexcerpt gleichzeitig ausgeführt, können Fehler wie z.B. eine fehlende Arbeitskopie vorkommen.

7.1 Suche und Extraktion mittels "SIARDexcerpt de.exe"-GUI

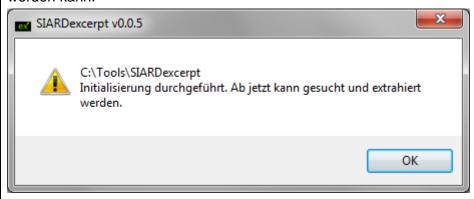


2 Initialisierung

- 1. Als erstes muss immer die Initialisierung durchgeführt werden.
- 2. Den Pfad zu SIARD Datei auswählen oder direkt eingeben.
- 3. Angaben zur Konfiguration eingeben
 - Den Pfad zu der Konfigurationsdatei auswählen oder direkt eingeben.
 - (..) wenn keine Vorgaben gemacht werden.
 - Tabellenname der gewünschten Haupttabelle
- 4. Initialisierung starten



Am Schluss der Initialisierung wird angezeigt, dass jetzt gesucht und extrahiert werden kann.



Hinweis:

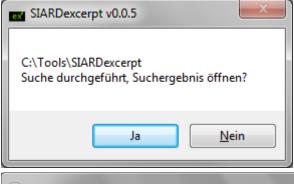
Bei der Initialisierung wird von jeder Konfiguration eine Kopie als "SIARDexcerpt.conf.xml" im Ordner "configuration" erstellt und alle "(..)" mit logischen Werten aus metadata.xml der SIARD-Datei erstellt. Es empfiehlt sich diese nach der Initialisierung zu Kopieren und unter einem anderen Namen abzuspeichern. Nachbesserungen in der Konfiguration können dann vorgenommen und beim nächsten Mal direkt verwendet werden.



- 1. Suche auswählen.
- Die gesuchten Werte eingeben. Dabei müssen die einzelnen Werte mit der Wildcard * getrennt werden. Die Reihenfolge wird übernommen. Leerschläge und andere Sonderzeichen müssen durch die Wildcard * ersetzt werden.
- Suche starten



Am Schluss der Suche wird angezeigt, dass die Suche durchgeführt wurde und das Ergebnis kann wenn gewünscht direkt angezeigt werden.

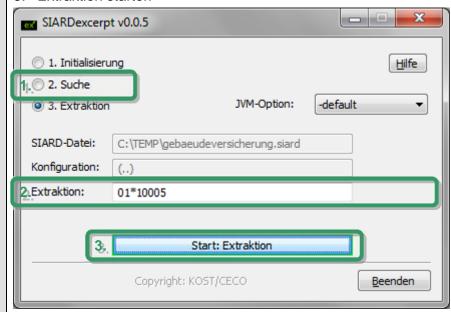




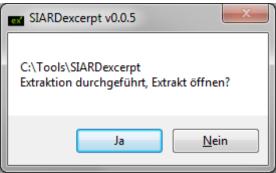
Anhand des Primärschlüssels, welcher in der Tabelle fett dargestellt wird, kann extrahiert werden.

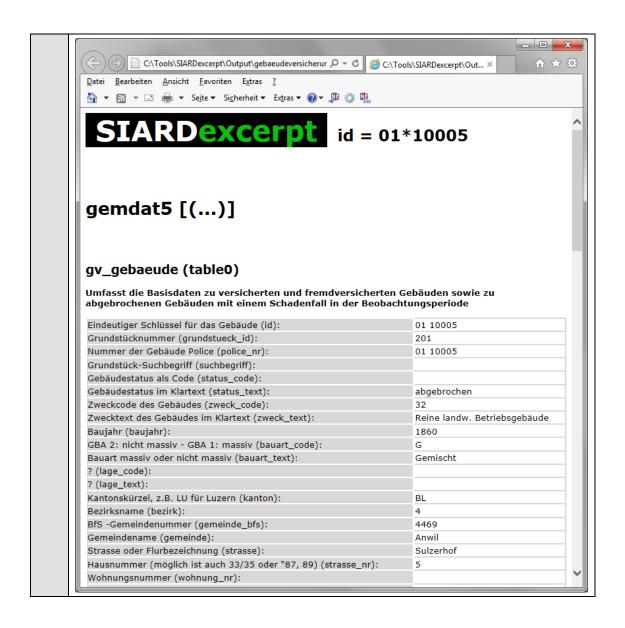
4 Extraktion

- 1. Extraktion auswählen.
- 2. Der Schlüssel eingeben. Leerschläge und andere Sonderzeichen müssen durch die Wildcard * ersetzt werden.
- 3. Extraktion starten



Am Schluss der Extraktion wird angezeigt, dass die Extraktion durchgeführt wurde und das Ergebnis kann wenn gewünscht direkt angezeigt werden.





7.2 Manuelle Suche und Extraktion

1 Eingabeaufforderung öffnen (Start → Ausführen ... → cmd [OK]) und in das gewünschte Arbeitsverzeichnis wechseln (CD C:\Tools\SIARDexcerpt)³.



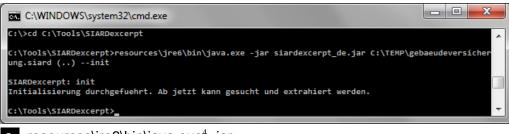
Allg. Anmerkungen:

Wenn ein Pfad Leerzeichen enthält, muss dieser in Anführungs- und Schlusszeichen eingegeben werden.

SIARDexcerpt kann auch von einem beliebigen Ort aus aufgerufen werden. Dies bedingt jedoch die Eingabe von absoluten Pfaden nicht nur im Programmaufruf selber, sondern auch in der Konfigurationsdatei.

2 Als erstes muss immer die Initialisierung durchgeführt werden.

SIARDexcerpt-Programmaufruf starten (die einzelnen Eingabebestandteile mit Leerzeichen trennen).



- A resources\jre6\bin\java.exe4 -jar
- B siardexcerpt_de.jar
- C:\TEMP\gebaeudeversicherung.siard
- D (..)
- ı --init
- A = Java-Befehl (resources\jre6\bin\java.exe -jar),
- B = Pfad und die Datei siardexcerpt_de.jar,
- c = Pfad und Name zur SIARD Datei
- **D** = Konfigurationsdatei oder (..) oder Name der Hauttabelle,
- **=** --init.

Die Initialisierung wurde durchgeführt, sobald der Text "Initialisierung durchgefuehrt" erscheint. Der *exit*-Status kann die folgenden Werte annehmen:

- 0 alles OK
- 1 Fehler im Programmaufruf
- 2 es sind Probleme bei der Initialisierung aufgetreten

Hinweis:

Bei der Initialisierung wird von jeder Konfiguration eine Kopie als "SIARDexcerpt.conf.xml" im Ordner "configuration" erstellt und die "(..)" mit logischen Werten aus metadata.xml der SIARD-Datei erstellt. Es empfiehlt sich diese nach der Initialisierung zu Kopieren und unter einem anderen Namen

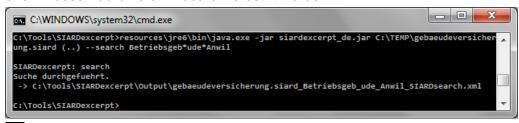
-

³ Das Laufwerk wird z.B. mit CD /D gewechselt.

⁴ Die Eingabe java – jar ist nur möglich, wenn Java Runtime Environment (JRE) Version 6 die Standardversion ist. Mit JRcE7 kann es zu Abbrüchen kommen. SIARDexcerpt ist bei Verwendung von JRE8 massiv langsamer.

abzuspeichern. Nachbesserungen in der Konfiguration können dann vorgenommen und beim nächsten Mal direkt verwendet werden.

Bei der Suche müssen die einzelnen Werte mit der Wildcard * getrennt werden. Die Reihenfolge wird übernommen. Leerschläge und andere Sonderzeichen müssen durch die Wildcard * ersetzt werden.



- A resources\jre6\bin\java.exe -jar
- B siardexcerpt_de.jar
- c C:\TEMP\gebaeudeversicherung.siard
- D (..)
- --search Betriebsgeb*ude*Anwil
- A = Java-Befehl (resources\jre6\bin\java.exe -jar),
- B = Pfad und die Datei siardexcerpt_de.jar,
- c = Pfad und Name zur SIARD Datei
- **D** = Konfigurationsdatei oder (..) oder Name der Hauttabelle,
- = --search sowie der Suchtext.

Die Suche wurde durchgeführt, sobald der Text "Suche durchgefuehrt" erscheint. Die Suchresultate sind in der ausgegebenen Datei (gebaeudeversicherung.siard_Betriebsgeb_ude_Anwil_SIARDsearch.xml) ersichtlich. Der *exit*-Status kann die folgenden Werte annehmen:

- 0 alles OK, mindestens eine Ergebnis wurde gefunden
- 1 Fehler im Programmaufruf
- 2 es sind Probleme bei der Suche aufgetreten

Bei der Extraktion müssen falls vorhanden Leerschläge und andere Sonderzeichen im Schlüssel durch die Wildcard * ersetzt werden.

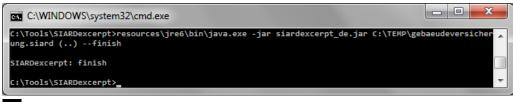


- A resources\jre6\bin\java.exe -jar
- B siardexcerpt_de.jar
- C:\TEMP\gebaeudeversicherung.siard
- D (..)
- --excerpt 01*10005
- A = Java-Befehl (resources\jre6\bin\java.exe -jar),
- B = Pfad und die Datei siardexcerpt de.jar,
- c = Pfad und Name zur SIARD Datei
- **D** = Konfigurationsdatei oder (..) oder Name der Hauttabelle,
- **E** = --excerpt sowie der Schlüssel.

Die Extraktion wurde durchgeführt, sobald der Text "Extraktion durchgefuehrt" erscheint. Der extrahierte Record ist in der ausgegebenen Datei (gebaeudeversicherung.siard_01_10005_SIARDexcerpt.xml) ersichtlich. Der *exit*-Status kann die folgenden Werte annehmen:

- 0 alles OK, Record konnte extrahiert werden
- 1 Fehler im Programmaufruf
- 2 es sind Probleme bei der Extraktion aufgetreten
- 5 Zum Schluss wird mit "--finish" alle temporären Dateien gelöscht.

Bei der Initialisierung wird von jeder Konfiguration eine Kopie als "SIARDexcerpt.conf.xml" im Ordner "configuration" erstellt und alle "(..)" mit logischen Werten aus metadata.xml der SIARD-Datei erstellt. Es empfiehlt sich diese vor dem Löschen zu Kopieren und unter einem anderen Namen abzuspeichern. Nachbesserungen in der Konfiguration können dann vorgenommen und beim nächsten Mal direkt verwendet werden.



- A resources\jre6\bin\java.exe -jar
- B siardexcerpt_de.jar
- c C:\TEMP\gebaeudeversicherung.siard
- D (..)
- --finish
- A = Java-Befehl (resources\jre6\bin\java.exe -jar),
- B = Pfad und die Datei siardexcerpt_de.jar,
- c = Pfad und Name zur SIARD Datei
- **D** = Konfigurationsdatei oder (..) oder Name der Hauttabelle,
- $\mathbf{E} = \text{--finish}.$

8 Urheberrecht

SIARDexcerpt ist eine Entwicklung der KOST. Alle Rechte liegen bei der KOST. SIARDexcerpt wurde im 2016 durch die KOST unter der GNU General Public License v3+ veröffentlicht.

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

SIARDexcerpt stützt sich auf folgende unveränderte Komponenten anderer Hersteller, welche direkt im Quellcode von SIARDexcerpt eingebunden sind:

Drittprogramm / -Komponente	Version	Lizenz
Apache Commons http://commons.apache.org/		Apache License 2.0
- commons-collections-3.2.1.jar	3.2.1	
- commons-configuration-1.6.jar	1.6	
- commons-digester-1.8.jar	1.8	
- commons-lang-2.4.jar	2.4	
- commons-logging-1.1.1.jar	1.1.1	
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 <u>http://sourceforge.net/projects/zip64file/</u>	1.02	GPL v2+ License

SIARDexcerpt stützt sich auf folgende unveränderte Komponenten anderer Hersteller, welche mit SIARDexcerpt abgegeben werden:

Drittprogramm / -Komponente		Version	Lizenz
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

Die Benutzer von SIARDexcerpt sind gehalten, die Lizenzbestimmungen all dieser Komponenten zu befolgen, welche im Verzeichnis SIARDexcerpt\license vorliegen.

9 Anhang

9.1 Beschreibung der Konfigurationsdatei

Die Konfigurationsdatei ist in verschiedenen Teilen aufgebaut, welche nachfolgend detailliert beschrieben werden.

9.1.1 Allgemeiner Teil

<pathtooutput>

Pfadangabe zum Outputverzeichnis von SIARDexcerpt. Der Initialwert ist **Output**. Dabei kann der absolute Pfad angegeben werden oder wie im Initialwert der relationale Pfad (aus Sicht von siardexcerpt_de.jar). Dieser Ordner muss nicht angelegt werden.

<pathtoworkdir>

Pfadangabe zum temporären Arbeitsverzeichnis von SIARDexcerpt. Der Initialwert ist **temp_SIARDexcerpt**. Dabei kann der absolute Pfad angegeben werden oder wie im Initialwert der relationale Pfad (aus Sicht von siardexcerpt_de.jar). Dieser Ordner muss nicht angelegt werden.

<pathtoxsl>

Pfadangabe zum Stylesheet der Extraktion. Der Initialwert ist (..). Dabei kann der absolute Pfad angegeben werden oder der relationale Pfad (aus Sicht von siardexcerpt_de.jar).

Bei (..) wird das XSL automatisch anhand der Anzahl Tabellen und deren Anzahl Spalten erstellt.

<pathtoxslsearch>

Pfadangabe zum Stylesheet der Suche. Der Initialwert ist **resources\SIARDexcerptSearch.xsl**. Dabei kann der absolute Pfad angegeben werden oder wie im Initialwert der relative Pfad (aus Sicht von siardexcerpt_de.jar).

<archive>

Name des Archivs. Der Initialwert ist **Staatsarchiv Zuerich**. Dieser wird in der Fusszeile beim Suchergebnis und beim extrahierten Record angezeigt.

9.1.2 maintable Teil

<mainfolder>

Nummer in welcher Tabelle gesucht werden soll respektive die Haupttabelle bei der Extraktion. Der Initialwert ist (...).

Bei (..) wird die Tabelle nach der Anzahl Primarykey ermittelt:

- 1. Tabelle mit dem einzigen Primarykey
- 2. Tabelle mit dem Primarykey, welcher am meisten verwendet wird
- 3. Wenn kein Primarykey wird die Tabelle mit den meisten Spalten eingesetzt

<mainname>

Name in welcher Tabelle gesucht werden soll respektive die Haupttabelle bei der Extraktion. Der Initialwert ist (...).

Bei (..) wird die Tabelle wie bei <mainfolder> ermittelt.

<title>

Titel der Suchergebnisse. Der Initialwert ist (..).

Bei (..) wird der Text "Suchergebnis aus der Tabelle <mainname>:" verwendet.

orimarykeyname>

Angabe wie der Hauptschlüssel bezeichnet werden soll. Der Initialwert ist (..). Bei (..) wird der Hauptschlüssel wie bei <mainfolder> ermittelt.

Angabe in welcher Zelle sich der Hauptschlüssel befindet. Der Initialwert ist (..). Bei (..) wird der Hauptschlüssel wie bei <mainfolder> ermittelt.

<cellname1> ... <cellname11>

Angabe wie die elf weiteren Spalten für die Ausgabe des Suchergebnisses bezeichnet werden sollen. Der Initialwert ist jeweils (...).

Bei (..) werden mit ersten 11 respektive 12 sinnvollen Spalten definiert und erstellt. Sinnvoll bedeutet, dass die verschiedenen types priorisiert werden:

- 1. Höchste Priorität (1) für "CHARACTER VARYING", "CHARACTER" und "DATE"
- 2. Priorität 2 für "DECIMAL", "NATIONAL CHARACTER VARYING" und "NATIONAL CHARACTER"
- 3. Priorität 3 für "BIGINT", "INTEGER", "SMALLINT" und "NUMERIC"
- 4. Priorität 4 Für "DOUBLE PRECISION", "FLOAT", "INTERVAL" und "REAL"
- 5. Priorität 5 für "TIME", "TIME WITH TIME ZONE", "TIMESTAMT" und "TIMESTAMP WITH TIME ZONE"
- 6. Priorität 6 für "BINARY VARYING", "BINARY", "BIT VARYING", "BIT" und "XML"
- 7. Priorität 7 für "BINARY LARGE OBJECT", "BOOLEAN", "CHARACTER LARGE OBJECT"
- 8. Bei Bedarf wird "NATIONAL CHARACTER LARGE OBJECT" und ggf andere verwendet

<cellnumber1> ... < cellnumber11>

Angabe in welcher Zelle sich die Spalten für die Ausgabe des Suchergebnisses befinden. Der Initialwert ist jeweils (..).

Bei (..) werden die Spaltennummern wie bei <cellname> ermittelt.

9.1.3 subtables Teil

<subtables>

Angaben⁵ über die Tabellen welche verknüpft werden sollen und den Schlüsseln. Der Initialwert ist (...).

Bei (..) wird für jede Tabelle <subtable>, welchen den primarykey als foreignkey anzieht, <name>, <folder> und <foreignkeycell> benötigt respektive ermittelt.

5

9.2 Funktionsprinzip SIARDexcerpt

