Ein Gemeinschaftsunternehmen von Schweizer Archiven



Anwendungshandbuch

Inhalt

1	Vorwort	
2	2 Systemvoraussetzungen	2
	B Offene Punkte / Feedback	
4	Installation	3
5	Konfiguration des SIARDexcerpt und der SIARD-Suche und Extrak	tion3
	5.1 Bestandteile der Konfigurationsdatei	3
6	Ressourcen von SIARDexcerpt	5
7	SIARDexcerpt verwenden	6
	7.1 Suche und Extraktion mittels «SIARDexcerpt.exe»-GUI	6
	7.2 Manuelle Suche und Extraktion	10
8	3 Urheberrecht	12
9	9 Anhang	13
	9.1 Beschreibung der Konfigurationsdatei	13
	9.2 Funktionsprinzip SIARDexcerpt	15

1 Vorwort

SIARDexcerpt ist eine java-basierte Anwendung zur Suche und Extraktion einzelner Records aus SIARD-Dateien¹. 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.

Initialisierung: 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 wie in Kapitel 9.1 beschrieben anhand metadata.xml automatisch ausgefüllt und als SIARDexcerpt.conf.xml temporär gespeichert.

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

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

Beenden: Zum Schluss werden die temporäre Konfiguration SIARDexcerpt.conf.xml sowie die entpackte SIARD-Datei wieder gelöscht.

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 8 respektive 1.8 inkl. JavaFX oder Liberica Full openJDK²

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.

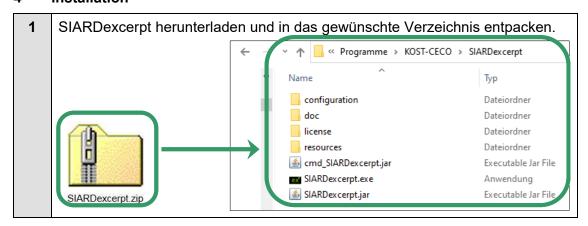
Bg/Km/Rc, 29.03.2021

¹ Die SIARD-Spezifikation (Software Independent Archiving of Relational Databases) kann von der eCH Website heruntergeladen werden:

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

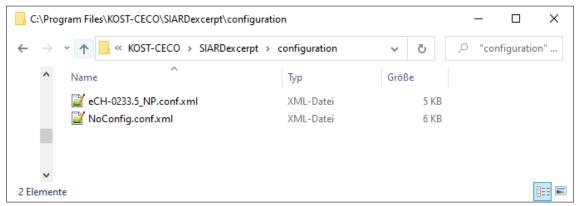
² Folgende Version wurde erfolgreich getestet: https://download.bellsw.com/java/14.0.2+13/bellsoft-jdk14.0.2+13-windows-amd64-full.msi

4 Installation



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 ++³ 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 «eCH-0233.5_NP.conf.xml» mitenthalten. Letztere dient für die Extraktion aus der Veranlagungsdatenbank respektive dem Veranlagungsauszug.

Bei der Initialisierung wird von jeder Konfiguration eine Kopie als «SIARDexcerpt.conf.xml» im Verzeichnis «USERHOME\.siardexcerpt\configuration» erstellt und alle «(..)» mit aus metadata.xml der SIARD-Datei abgeleiteten Werten 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.

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.

SIARDexcerpt_Anwendungshandbuch_v1.0.0.docx Bg/Km/Rc, 29.03.2021

³ Die Portable Notepad ++ Applikation kann unter http://portableapps.com/de/apps/development/notepadpp portable heruntergeladen werden.

5.1.1 Allgemeiner Teil

xml-Tag	Beschreibung: Initialwert
<pathtoxsl></pathtoxsl>	Pfadangabe zum Stylesheet der Extraktion. In unserem Beispiel ist dies: ()
<pathtoxslsearch></pathtoxslsearch>	Pfadangabe zum Stylesheet der Suche: resources\SIARDexcerptSearch.xsl
<archive></archive>	Name des Archivs: Archiv
<insensitive></insensitive>	Gross- und Kleinschreibung ignorieren: yes
<sed></sed>	Angabe ob mit sed alle einzelnen Daten der row auf eine Zeile gebracht werden soll: no

5.1.2 maintable Teil

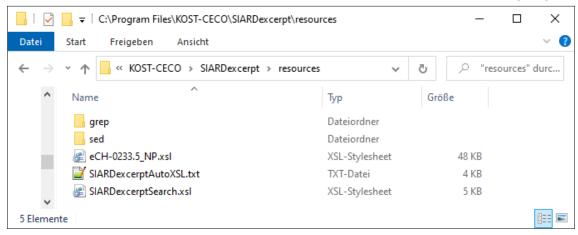
xml-Tag	Beschreibung: Beispiel	
<mfolder></mfolder>	Ordner, in welcher Tabelle gesucht werden soll, respektive die Haupttabelle bei der Extraktion: ()	
<mname></mname>	Name der Tabelle, in welcher gesucht werden soll, respektive die Haupttabelle bei der Extraktion: ()	
<mtitle></mtitle>	Titel der Suchergebnisse: ()	
<mpkname></mpkname>	Angabe, wie der Hauptschlüssel bezeichnet werden soll: ()	
<mpkcell></mpkcell>	Angabe, in welcher Zelle sich der Hauptschlüssel befindet: ()	
<mschemaname></mschemaname>	Schemaname in welcher sich die Haupttabelle befindet: ()	
<mschemafolder></mschemafolder>	Schemaordner in welcher sich die Haupttabelle befindet: ()	
<mc1name> <mc11name></mc11name></mc1name>	Angabe, wie die weiteren Spalten für die Ausgabe des Suchergebnisses bezeichnet werden sollen: ()	
<mc1number> <mc11number></mc11number></mc1number>	Angabe, in welcher Zelle sich die Spalten für die Ausgabe des Suchergebnisses befindet: ()	

5.1.3 subtables Teil

xml-Tag	Beschreibung: Initialwert
<st1keyname> <st1name></st1name></st1keyname>	Angaben ⁴ über die Tabellen, welche verknüpft werden sollen, und die Schlüssel: ()
<st1folder></st1folder>	Es können bis zu 20 Tabellen verknüpft werden.
<st1fkcell></st1fkcell>	
<st1schemafolder></st1schemafolder>	
<st1schemaname></st1schemaname>	

6 Ressourcen von SIARDexcerpt

Sämtliche Ressourcen von SIARDexcerpt sind im Unterordner «resources» abgelegt.



«grep» und «sed» 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. «eCH-0233.5_NP.xsl" ermöglicht die Darstellung des einzelnen Veranlagungsauszug-Record in Form einer vereinfachten Form einer Steuererklärung.

-

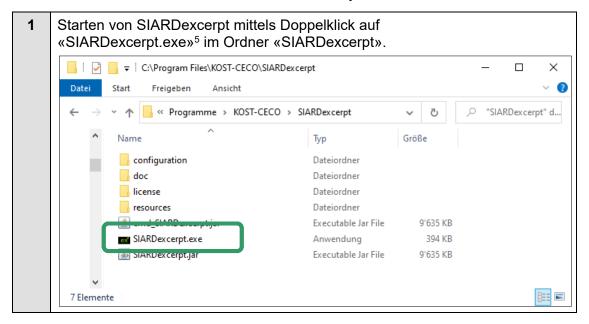
⁴ Pro <subtable> wird <name>, <folder>, <foreignkeycell>, <schemafolder> und <schemaname> benötigt respektive ermittelt.

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.

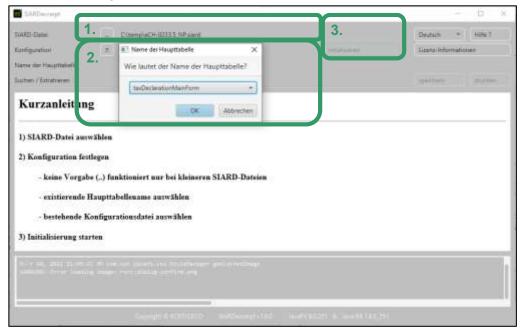
7.1 Suche und Extraktion mittels «SIARDexcerpt.exe»-GUI



-

⁵ Wird openJDK² verwendet, muss das GUI via «SIARDexcerpt.jar» und «Öffnen mit...» gestartet werden. Dabei muss das java.exe von openJDK angeben werden.

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



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



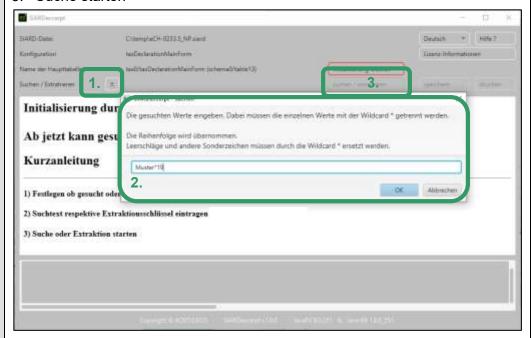
Hinweis:

Bei der Initialisierung wird von jeder Konfiguration eine Kopie als «SIARDexcerpt.conf.xml» im Verzeichnis

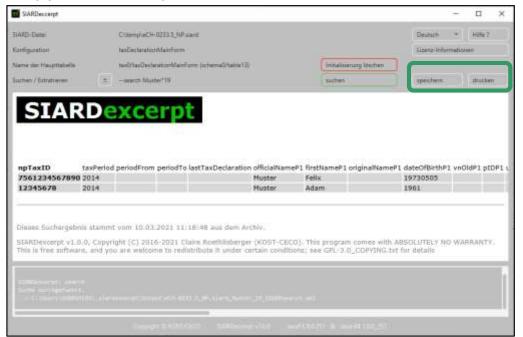
«USERHOME/.siardexcerpt/configuration» erstellt und alle «(..)» mit aus metadata.xml der SIARD-Datei abgeleiteten Werten erstellt. Es empfiehlt sich diese nach der Initialisierung zu Kopieren und unter einem anderen Namen abzuspeichern. Nachbesserungen in der Konfiguration können dann vorgenommen und beim nächsten Mal direkt verwendet werden.

3 Suche

- 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.
- 3. Suche starten



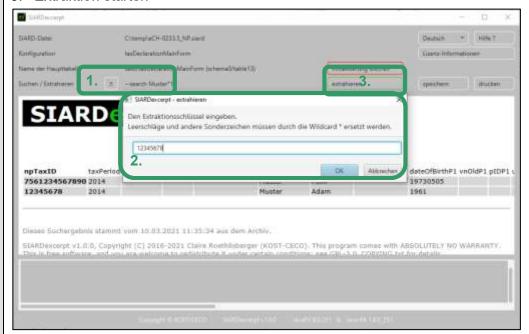
Am Schluss der Suche wird angezeigt, dass die Suche durchgeführt wurde und das Ergebnis wird angezeigt. Wenn gewünscht kann das Ergebnis anderweitig gespeichert oder gedruckt werden.



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

4 Extraktion

- 1. Extraktion auswählen.
- 2. Den 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 wird angezeigt. Wenn gewünscht kann das Ergebnis anderweitig gespeichert oder gedruckt werden.



7.2 Manuelle Suche und Extraktion

1 Eingabeaufforderung öffnen (Start → Ausführen ... → cmd [OK]) und in das gewünschte Arbeitsverzeichnis wechseln (CD «C:\Program Files\KOST-CECO\SIARDexcerpt»). Aufbau SIARDexcerpt Befehl:

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.

Als erstes muss immer die Initialisierung durchgeführt werden.

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

```
C:\Program Files\KOST-CECO\SIARDexcerpt>java -jar cmd_SIARDexcerpt.jar c:\temp\eCH-0233.S_NP.siard configuration\eCH-0233.S_NP.conf.xml --de --init

SIARDexcerpt: init
Initialisierung durchgefuehrt. Ab jetzt kann gesucht und extrahlert werden.

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

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

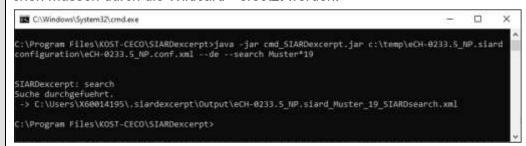
- 0 alles OK
- 1 Fehler im Programmaufruf
- 2 es sind Probleme bei der Initialisierung aufgetreten

Hinweis:

Bei der Initialisierung wird von jeder Konfiguration eine Kopie als «SIARDexcerpt.conf.xml» im Verzeichnis

«USERHOME/.siardexcerpt/configuration» erstellt und alle «(..)» mit aus metadata.xml der SIARD-Datei abgeleiteten Werten 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.

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.



Die Suche ist beendet, sobald der Text "Suche durchgefuehrt" erscheint. Die Suchresultate sind in der ausgegebenen Datei (eCH-0233.5_NP.siard_Muster_19_SIARDsearch.xml) ersichtlich. Der *exit*-Status kann die folgenden Werte annehmen:

- 0 alles OK, mindestens ein Ergebnis wurde gefunden
- 1 Fehler im Programmaufruf
- 2 es sind Probleme bei der Suche aufgetreten
- **4** Bei der Extraktion müssen falls vorhanden Leerschläge und andere Sonderzeichen im Schlüssel durch die Wildcard * ersetzt werden.

```
C:\Users\X60814195\.siardexcerpt\Output\eCH-0233.5_NP.siard_Muster_19_SIARDsearch.xml

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 --de --excerpt 12345678

SIARDexcerpt: extract
Extraktion durchgefuehrt.
-> C:\Users\X60814195\,siardexcerpt\Output\eCH-0233,5_NP.siard_12345678_SIARDexcerpt.xml

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

Die Extraktion ist beendet, sobald der Text "Extraktion durchgefuehrt" erscheint. Der extrahierte Record ist in der ausgegebenen Datei (eCH-0233.5_NP.siard_12345678_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 werden mit "--finish» alle temporären Dateien gelöscht.

Bei der Initialisierung wird von jeder Konfiguration eine Kopie als «SIARDexcerpt.conf.xml» im Verzeichnis

«USERHOME/.siardexcerpt/configuration» erstellt und alle «(..)» mit aus metadata.xml der SIARD-Datei abgeleiteten Werten 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.



SIARDexcerpt_Anwendungshandbuch_v1.0.0.docx Bg/Km/Rc, 29.03.2021

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		n Lizenz
Apache Commons http://commons.a	apache.org/	Apache License 2.0
- commons-io-2.6.jar - commons-logging-1.2.jar	2.6 1.2	
Apache log4j http://logging.apach	<u>e.org/log4j/</u> 1.2.12	Apache License 2.0
Apache Xerces http://xerces.a	apache.org/ 2.7.1	Apache License 2.0
Junit 4.12 <a documents"="" href="http://www.ntp://ww.ntp://ww.ntp://ww.ntp://ww.ntp://ww.ntp://ww</th><th>w.junit.org/ 4.12</th><th>CPL v1.0</th></tr><tr><th>Spring Framework API http://static.springsource.org/spring/documents	<u>cs/3.0.x/api/</u> 5.0.8	Apache License 2.0
zip64 http://sourceforge.net/project	ts/zip64file/ 1.02	GPL v2+ License

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

Drittprogramm / -Komponente		Version	Lizenz
GNU grep	http://www.gnu.org/software/grep	2.4.2	GPL v3+ License
GNU sed	https://www.gnu.org/software/sed	4.4	GPL v3+ 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

<pathtoxsl>

Pfadangabe zum Stylesheet der Extraktion. Der Initialwert ist

(..). Dabei kann der absolute Pfad angegeben werden oder der relative Pfad (aus Sicht von cmd_SIARDexcerpt.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 **Archiv**. Dieser wird in der Fusszeile beim Suchergebnis und beim extrahierten Record angezeigt.

<insensitive>

Gross- und Kleinschreibung bei der Suche ignorieren. Der Initialwert ist yes.

<sed>

Angabe ob mit sed alle einzelnen Daten der row auf eine Zeile gebracht werden soll. Der Initialwert ist **no**. Die meisten SIARD-Programme schreiben bereits alle Daten der row auf einer Zeile.

9.1.2 maintable Teil

<mfolder>

Nummer der Tabelle, in welcher gesucht werden soll, respektive der 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

<mname>

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

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

<mtitle>

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

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

<mpkname>

Angabe, wie der Hauptschlüssel bezeichnet werden soll. Der Initialwert ist (..).

<mpkcell>

Angabe, in welcher Zelle sich der Hauptschlüssel befindet. Der Initialwert ist (..).

<mschemaname>

Schemaname in welcher sich die Haupttabelle befindet. Der Initialwert ist (..).

<mschemafolder>

Schemaordner in welcher sich die Haupttabelle befindet. Der Initialwert ist (..).

<mc1name> ... <mc11name>

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

Bei (..) werden die Spalten gemäss ihrem Datentyp wie folgt priorisiert und die ersten 11 respektive 12 Spalten ausgewählt:

- 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

<mc1number> ... <mc11number>

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

9.1.3 subtables Teil

<subtables>

Angaben⁶ über die Tabellen, welche verknüpft werden sollen, und die Schlüssel. Es können bis zu 20 Tabellen verknüpft werden. Der Initialwert ist (..).

Bei (..) wird für jede Tabelle, welche den primarykey als foreignkey referenziert, die Angaben ermittelt.

SIARDexcerpt_Anwendungshandbuch_v1.0.0.docx Bg/Km/Rc, 29.03.2021

⁶ <st1keyname>, <st1name>, <st1folder>, <st1fkcell>, <st1schemafolder>, <st1schemaname>

9.2 Funktionsprinzip SIARDexcerpt

