



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

Generátor parserů pro XML code-snippets

BI-SP2 Uživatelská příručka

Vedoucí týmu: Jan Lejnar

Členové : Filip Šmíd, Luis Sanchez, Jakub Tkáč, Jakub Šedý, Matěj Schuh

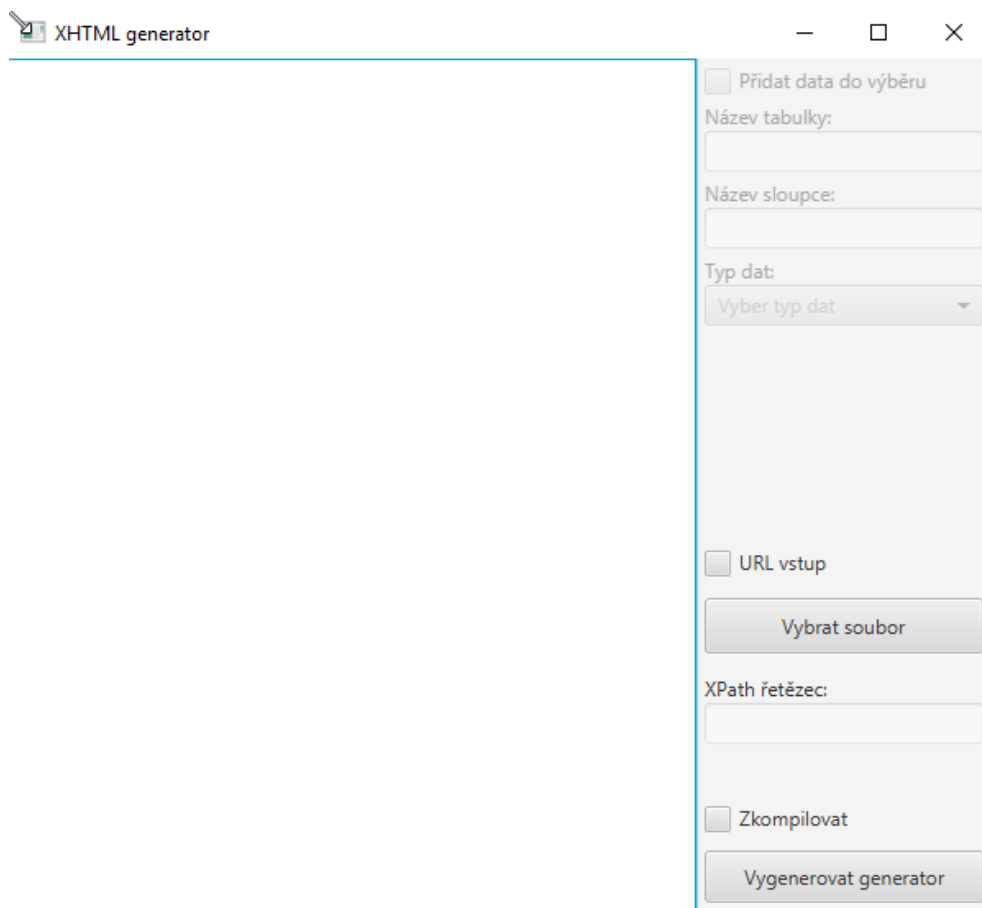
Vedoucí práce: Ing. Michal Valenta, Ph.D.

28. ledna 2017

Obsah

1	Hlavní obrazovka	2
2	Popis vygenerování parseru krok po kroku	3
2.1	Zacházení s nezbuilděným generátorem	4
2.2	Spuštění generátoru přes CLI	6
2.3	Možnost stažení xml souboru z webu	7
3	Uložení výsledků do DB	7
4	Obsah databáze	10

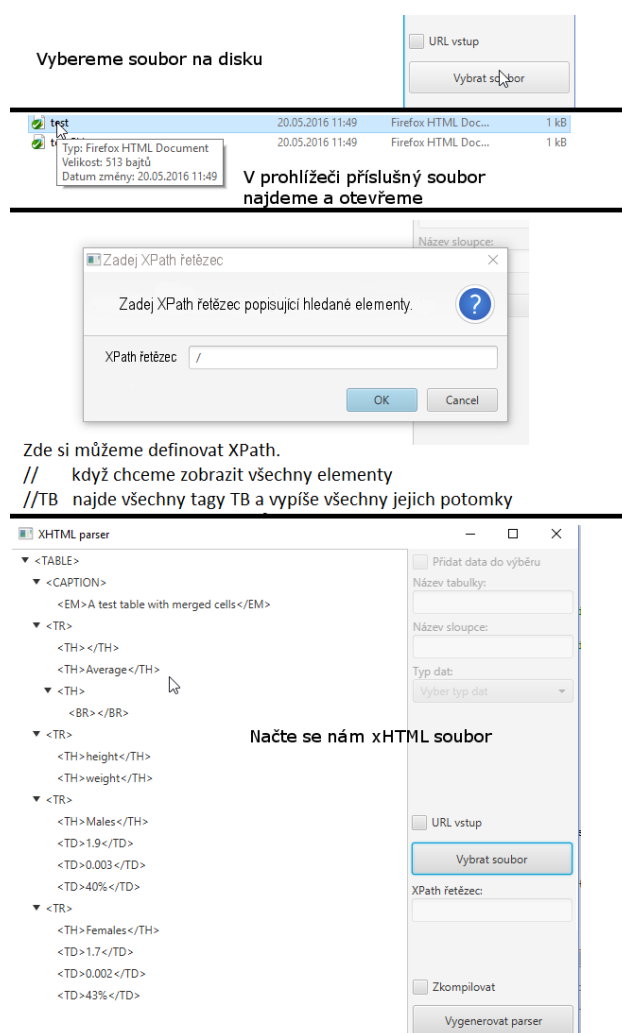
1 Hlavní obrazovka



Po spuštění programu můžeme vidět podobnou hlavní obrazovku, která slouží k ovládání celé aplikace. Nachází se zde možnosti výběru *xml souboru*, stažení *xml souboru* přímo z webu. Dále se zde nachází základní ovládací panely k vybraným datům z kódu. Nakonec se zde nachází možnost vytvoření samotného generátoru parseru.

2 Popis vygenerování parseru krok po kroku

Klikneme na tlačítko *Vybrat soubor*, a následně vybereme náš soubor. Poté máme možnost zadat XPath. Pokud chceme vybrat všechny elementy tak zadáme `//`. Pokud chceme kontrétní tak můžeme zvolit `//TB`, které najde všechny výskyty tagu TB v dokumentu a zobrazí všechny jeho potomky. Případně i kombinace se zanořením `//TB/knihy` zobrazí vše vyhovující posloupnosti tagů TB a potomka knihy.



Až máme soubor, tak klikneme na zajímavé části v kódu. Následně vpravo navolíme název tabulky, sloupce, typ dat a přidáme do výběru. Pokud zapomeneme přidat do výběru, tak generátor dané data do parseru nepřidá.

Až budeme mít jednotlivé položky přidané, klikneme na *Vygenerovat generátor*. Máme zde možnost buď rovnou vygenerovat binárku `<nazevParseru>.jar` a nebo vygenerovat parser, který později bude potřeba zbuildit.

The screenshot shows the 'Vygenerovat generátor' (Generate generator) window. It is divided into three main sections:

- Top Section:** A tree view on the left shows the XML structure of the generated parser. The selected item is `A test table with merged cells`. The right panel shows options for adding data to the selection: ☐ Přidat data do výběru, a text input for 'Název tabulky:' (Table name), a text input for 'Název sloupce:' (Column name), and a dropdown for 'Typ dat:' (Data type) with the option 'Vyber typ dat'.
- Middle Section:** The tree view shows the next step, where the 'caption' and 'column' attributes are being filled in. The right panel shows ☒ Přidat data do výběru, the 'Název tabulky:' set to 'tabulka', 'Název sloupce:' set to 'captionSloupec', and 'Typ dat:' set to 'text'.
- Bottom Section:** The tree view shows the final step, where the data is being compiled. The right panel shows ☒ Zkompilovat and a button 'Vygenerovat parser'.

Below the tree view, there is a text box with the instruction: 'Až vybereme všechny zajímavé části, tak dáme vygenerovat parser. Je možnost ho rovnou i zkompilovat.' (When we select all interesting parts, we will generate the parser. There is an option to compile it right away.)

At the bottom, there is a section 'A uložíme' (Where we save it) with a text input for 'Název souboru:' (File name) set to 'mujparser' and a dropdown for 'Uložit jako typ:' (Save as type) set to 'pXHTML'. There are buttons 'Uložit' (Save) and 'Zrušit' (Cancel).

Below this, there is a 'Hotovo' (Done) dialog box with the message 'Parser byl vygenerován.' (Parser was generated.) and an 'OK' button.

2.1 Zacházení s nezbuilděným generátorem

Pokud jsme nezaškrtnuli volbu *Zkompilovat* a následně klikli na *Vygenerovat generátor*, tak budeme si muset program sami zkompilovat. Prvně si vybereme, kam uložíme soubory potřebné ke kompilaci.

The screenshot shows a file save dialog box. It has a text input for 'Název souboru:' (File name) set to 'generated' and a dropdown for 'Uložit jako typ:' (Save as type). There are buttons 'Uložit' (Save) and 'Zrušit' (Cancel).

Situace ve složce bude následující.

[.]	<DIR>	20.12.2016 19:58	—
[generated]	<DIR>	20.12.2016 19:58	—
[.]	<DIR>	20.12.2016 19:58	—
[gradle]	<DIR>	20.12.2016 19:58	—
[libs]	<DIR>	20.12.2016 19:58	—
[parsemodel]	<DIR>	20.12.2016 19:58	—
[resources]	<DIR>	20.12.2016 19:58	—
[src]	<DIR>	20.12.2016 19:58	—
settings	gradle	51	20.12.2016 19:58 -a-
README	md	11	20.12.2016 19:58 -a-
gradlew	bat	2 176	20.12.2016 19:58 -a-
gradlew		5 242	20.12.2016 19:58 -a-
build	gradle	809	20.12.2016 19:58 -a-

Přes příkazový řádek vejdemo do složky a spustíme následující příkaz

Pro Windows:

gradlew.bat createJar

Pro linux:

pokud potřeba, tak přidáme práva

chmod +x ./gradlew

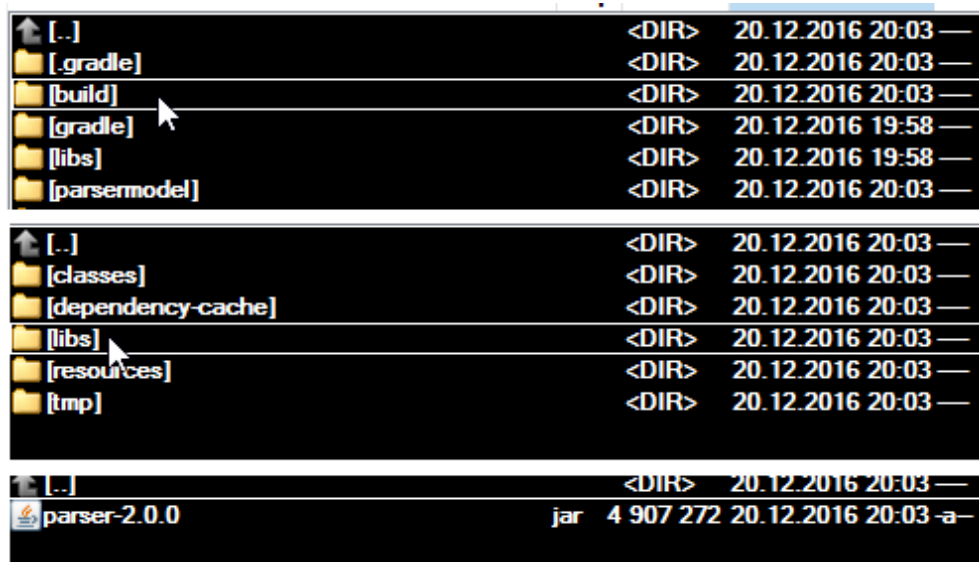
./gradlew createJar

```
C:\Users\SiRonnie\Parserr\generated>gradlew.bat createJar
Starting a Gradle Daemon (subsequent builds will be faster)
Download https://repo1.maven.org/maven2/com/google/code/findbugs/jsr305/3.0.1/jsr305-3.0.1.pom
Download https://repo1.maven.org/maven2/org/sonatype/oss/oss-parent/7/oss-parent-7.pom
Download https://repo1.maven.org/maven2/com/google/code/findbugs/jsr305/3.0.1/jsr305-3.0.1.jar
:parsemodel:compileJava
warning: [options] bootstrap class path not set in conjunction with -source 1.7
Note: C:\Users\SiRonnie\Parserr\generated\parsemodel\src\main\java\parsemodel\tree\foundData\FoundData.java uses unchecked or unsafe operations.
Note: Recompile with -Xlint:unchecked for details.
1 warning
:parsemodel:processResources UP-TO-DATE
:parsemodel:classes
:parsemodel:jar
:compileJava
warning: [options] bootstrap class path not set in conjunction with -source 1.7
1 warning
:processResources
:classes
:createJar

BUILD SUCCESSFUL
Total time: 20.404 secs
```

To nám zkompiluje generátor a vytvoří výsledný spustitelný Parser.

Parser najdeme ve složce *generated/build/libs/parser-2.0.0.jar*.



2.2 Spuštění generátoru přes CLI

Je zde také možnost obejít GUI a vygenerovat parser přes CLI. Pro toto generování ale potřebujeme speciální předpřipravené soubory. Postup spuštění na obrázku níže.

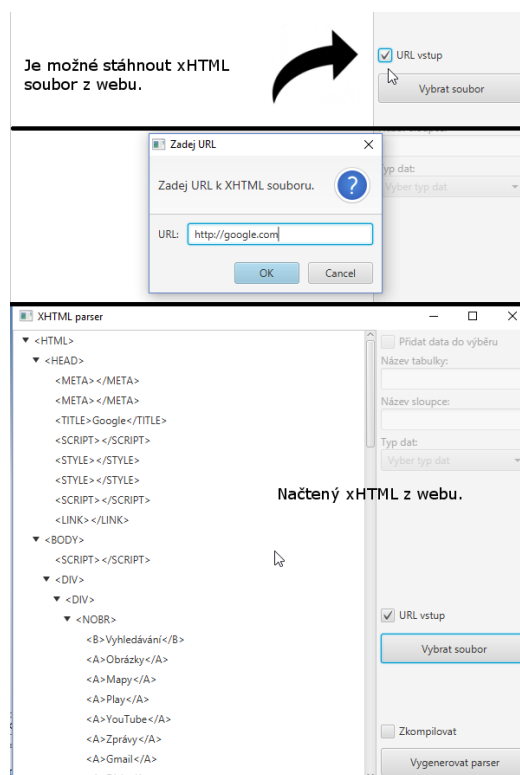
```
\parserWorkingDir>"%JDK_HOME%\bin/java.exe" -jar Parser.jar -nogui -compile in.xml out.jar
Parser byl vygenerován.
```

Příklad spuštění generátoru parseru z CLI.

Překládáme pomocí java.exe -jar (lze najít v jdk{verze}/bin/java.exe)
Parser.jar - samotný program pro generování parseru (NBProject/dist/Parser.java)
-nogui - argument, který udává, že nechceme spustit v GUI
-compile - pro kompilaci
in.xml - vstupní soubor s již předpřipravenými daty (ukázka v NBProject/testCLI.xhtml)
out.jar - takto se bude jmenovat zkompileovaná binárka parseru

2.3 Možnost stažení xml souboru z webu

Při výběru souboru je možnost zadat url adresu a soubor stáhnout z webu. Pro tuto možnost klikneme před výběrem souboru na check-box *URL vstup* a následně na *Vybrat soubor*. Do políčka zadáme URL adresu.



3 Uložení výsledků do DB

V místě, kam jsme vygenerovali generátor přes příkazový řádek můžeme uložit data do databáze.

- Pokud si nevíme rady, tak přes `java -jar Parser.jar -help` si můžeme vypsát nápovědu.


```

C:\Users\SiRonnie>java -jar Pars.jar -help
usage: Two options:
[1] parser.jar -cs <connection string> -db <name> -f <file> [-s <period>]
[2] parser.jar -cs <connection string> -db <name> -u <url> [-s <period>]

-cs,--cString <string>    Connection string for database. For more
                           information look at our User manual.
-db,--dbName <db>         Name of our database.
-f,--file <file>          If we are using file from disc. Give here
                           relative path to it.
-help                     Basic usage.
-s,--service <period>    For start our program as service. For more
                           information look at our User manual.
-u,--url <url>            If we are using xml file from web. Assign it
                           with https://www.
-version                  Version of our program.

```

Príkaz pro spustění parseru je následující.

```

C:\Users\SiRonnie>java -jar xhtmlParser1.jar -cs sp_16_valenta/SkoL.!@oracle.fit.cvut.cz:1521/ORACLE -db Oracle -u https://www.google.cz

```

Príkaz se skládá z:

- java -jar
- <Parser.jar> náš vygenerovaný generátor
- -f <file> xml soubor, který jsme dávali do generátoru
- -db <dbName> typ databáze, v základu předpřipravené
 - Oracle
 - MySQL
 - Postgres
- -cs <connectionString> náš connection string do databáze ve tvaru

<uzivatelske jmeno>/<heslo>@<host>:<port>/<JMENO DATABAZE>
- -s <cislo[s/S,m/M,h/H,d/D]> poběží náš program jako služba, kde

číslo udává jednotku následovanou jednotku času.
 - s/S sekunda
 - m/M minuta
 - h/H hodina

– d/D den

ukázka: -s 10s nám bude ukládat do databáze data každých 10 sekund

- -u <url> pokud nebudeme používat soubor, ale parsovat webovou stránku, tak ji zde zadáme včetně https://www.
- -help ukáže nám nápovědu, jaké argumenty lze zadat
- -version verze našeho programu

Ukázky:

Se souborem uloženým na disku.

```
java -jar <NášParser>.jar -f <ParsovanySoubor>.xml -db <typ databaze>
-cs <connectionString>(connectionString ve tvaru <uzivatelske jmeno>/
<heslo>@<host>:<port>/<JMENO DATABAZE>) -s <CisloHodnota> (Vo-
litelne, pokud chceme aby to behalo jako sluzba.)
```

Se souborem z webu.

```
java -jar <NášParser>.jar -u <ParsovanaStranka> -db <typ databaze>
-cs <connectionString>(connectionString ve tvaru <uzivatelske jmeno>/
<heslo>@<host>:<port>/<JMENO DATABAZE>) -s <CisloHodnota> (Vo-
litelne, pokud chceme aby to behalo jako sluzba.)
```

Příkaz bude vypadat následovně.

```
java -jar Parser.jar -u http://www.google.cz -db Oracle -cs uzivatelskeJme-
no/heslo@oracle:1521/ORACLE -s 10s
```

Aplikace podporuje Oracle, MySQL a PostgreSQL databáze. Pro přidání podpory další databáze navštivte prosím naši Instalační a programátorskou příručku.

4 Obsah databáze

The screenshot displays the Oracle SQL Developer interface. At the top, two query result windows are visible: 'FEMALE' and 'MALE'. The 'FEMALE' window shows a single row with values: ID=1, AVGHEIGHT=1.7, AVGWEIGHT=0.002, REDEYESPROBAB=30. The 'MALE' window shows a single row with values: ID=1, AVGHEIGHT=1.9, AVGWEIGHT=0.003, REDEYESPROBAB=40. Below these, a third window shows the schema details for the 'MALE' table. The schema table has columns: COLUMN_NAME, DATA_TYPE, NULLABLE, DATA_DEFAULT, COLUMN_ID, and COMMENTS. The table structure is as follows:

COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
ID	NUMBER (38, 0)	No	(null)	1	(null)
AVGHEIGHT	FLOAT	Yes	(null)	2	(null)
AVGWEIGHT	FLOAT	Yes	(null)	3	(null)
REDEYESPROBAB	NUMBER (38, 0)	Yes	(null)	4	(null)

Below the schema table, the text 'Výsledek v databázi' is displayed.

Po provedení předchozího kroku můžeme v naší databázi vidět naše vybrané data.