



FACULTY
OF INFORMATICS

Masaryk University

Searching Open Document Spreadsheets

.....

Java Implementation, Team B

M. Majer, S. Markosová, V. Novotný, and P. Kratochvíla

Analysis

The Open Document Format specification

ODS documents come in two flavours

1. XML documents and
2. ZIP archives containing the file content.xml.

and are usually extended.

- The spec permits the inclusion of arbitrary elements and attributes beyond what is specified by the Relax NG schema.

Holes in tables are represented by a single cell with the number-rows and columns-repeated attributes.

- This makes it difficult to tell the coordinates of a cell without pre-processing the entire document beforehand.

Design

The high-level overview

There are three packages.

- **common** – The effective code for processing ODS documents,
 - Assigned to Matěj Majer and Svetlana Markosová.
- **cli** – The command-line interface of the program,
 - Assigned to Petr Kratochvíla.
- **gui** – The graphical user interface of the program.
 - Assigned to Vít Novotný.

The common package

The first draft

1. Deserialize the input XML files to a tree structure of Java objects that correspond to the pruned DOM tree with coordinates added to the cells.
2. Perform any further queries on this internal java object representation.

Straightforward, but makes sparse use of XML technologies.
Therefore, this draft was scrapped.

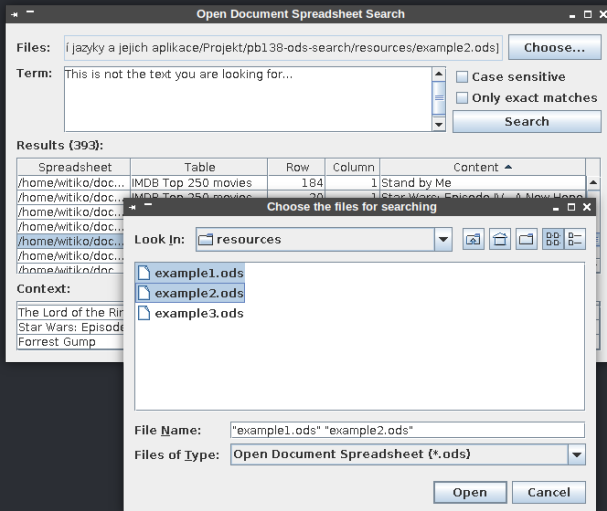
The common package

The second draft

1. Convert the input XML files to an intermediary XML format with resolved cell coordinates via a XSLT transformation.
2. Unlike the extended ODS documents, this intermediary format can now be validated against a XML schema.
3. Perform any further queries on this internal XML representation via XPath.

Extensive use of XML technologies. This draft was implemented.

The gui package



The cli package

```
usage: java -jar pb138-ods-search.jar  
        [-h] [-i] [-x] -s <arg> [FILE]...
```

-h,--help	Help
-i,--ignore-case	Use case-insensitive search
-s,--string <arg>	String to be used for search
-x,--exact-match	Use exact match search