

PREPARATION

A. Hardware, Software and Data Storage for Students

Hardware:

- Every student brings her/his laptop to the ISP (“bring your own device”)

Software - Students will need already installed software on each laptop:

- **Excel** (or other spreadsheet software, able to read Excel files)
- **Rapidminer Studio** (<https://rapidminer.com/products/studio>)
 - Community Edition is free (up to 10000 data rows, more than enough) or Academic License to students, professors and researchers is also free upon registration
 - together with the „Text Processing“ Extension (also free)
 - some “processes” (= programming code) for DIPCAT provided in advance
 - ***see D. for detailed instructions!***
- **recommended: Notepad++** (instead of Editor.exe), Open Source
 - <https://notepad-plus-plus.org/>

Data storage convention:

- Create a subfolder on your laptop named “DIPCAT” to store everything that you will need for the case.

B. Legal Framework

Directives, regulations, auditing standards etc. will not be printed, but are available as pdfs.

C. Pre-reading Assignments for Students

Students will need to read in preparation for the text mining parts of the ISP:

- Regular Expressions (RegEx) Quick Start.pdf

D. Instructions for Rapidminer

1. Installation of Rapidminer Studio Software

- Install Rapidminer Studio from:
<https://rapidminer.com/products/studio>
 - Choose “Download” and “Educational purposes”
- Installation guide:
<https://docs.rapidminer.com/latest/studio/installation/>

2. Tutorials/Learning – Alternatives:

- From the popping window “Welcome to RapidMiner Studio!”
 - select “Learn” tab
 - then “Get Started” → 1. Operators and Processes”

or
- From the top menu: “Help”
 - select “Tutorials”
 - From the popping window “Welcome to RapidMiner Studio!” select “Learn” tab
 - then “Get Started” → 1. Operators and Processes”

or
- Go to: Academy RapidMiner
 - <https://academy.rapidminer.com/learning-paths/get-started-with-rapidminer-and-machine-learning>
 - watch <https://academy.rapidminer.com/learn/video/rapidminer-studio-gui-intro>

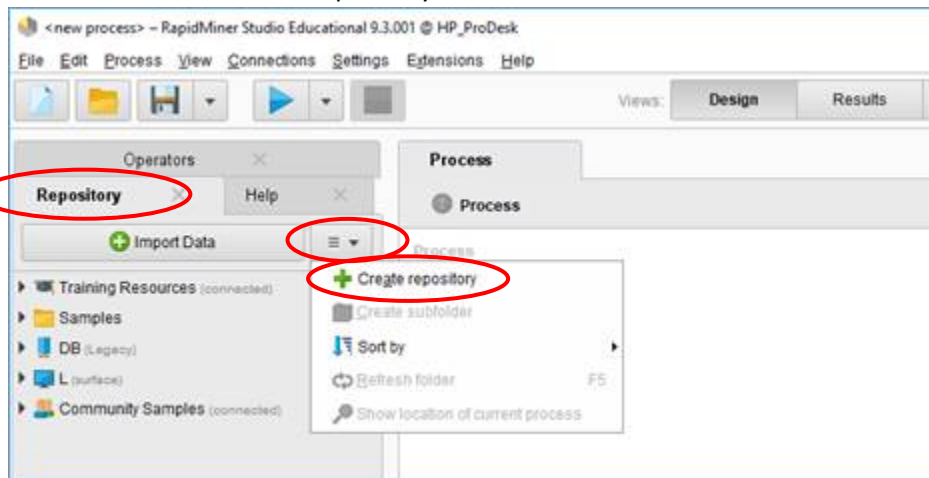
3. Installation of „Text Processing“ Extension

- Guide for adding extensions:
<https://docs.rapidminer.com/latest/studio/installation/adding-extensions.html>

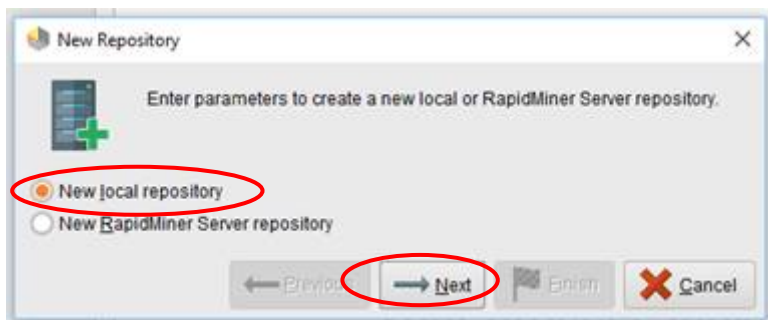
or
- Video <https://academy.rapidminer.com/learn/video/the-rapidminer-marketplace-extensions>
- → select „Text Processing“ Extension!

4. Prepare Data and Processes

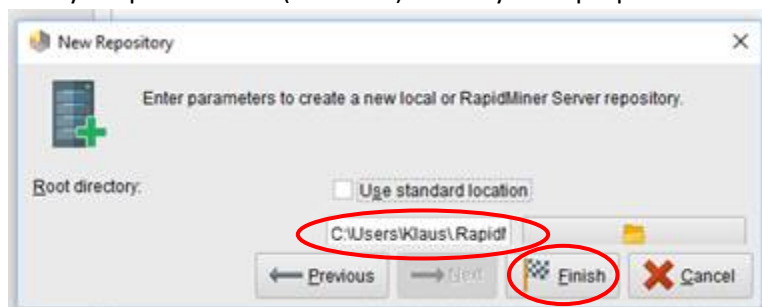
- a) Take the subfolder “RePoDIPCAT” and copy it into the local folder “DIPCAT” on your laptop
- this contains program code files that are called “processes” in RapidMiner
 - it also contains data for cases “H. Document Analysis” [contracts] and “I. Consistency of Narratives” [reports]
- b) You must tell the computer where it can find the process files (and data). To do this create a new “Repository”



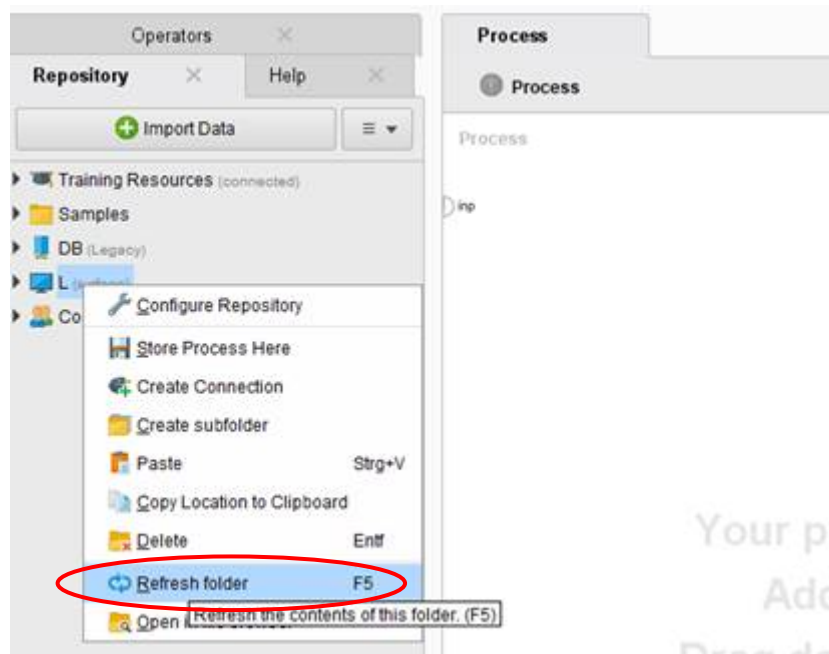
- c) Make it a new LOCAL repository (NOT in the cloud):



- d) Do NOT click “Use standard location”, INSTEAD tell the computer where the subdirectory “RePoDIPCAT” with your process files (and data) are on your laptop:



- e) You might need to “Refresh Folder” with a click-right (or F5) on your new Repository to make files appear (or restart Rapidminer):



5) ATTENTION LATER in the course:

- Double-click on the processes opens them.
- Run the process by pressing the play symbol (little triangle).
- ATTENTION!
In the parameters of the process “10 Extract the Contract Data (Task)”
you have to change the location of the directory where you stored the contract files.
On your computer, it would be *somewhere/DIPCAT/RepoDIPCAT/contracts*.
Otherwise, you will get an error message.
→ Please also see instructions in the yellow fields of the process code.
- ATTENTION!
The same applies to the process “32 Results for a WordList (negative)” etc.
You have to change the location of the directory where you stored the contract files.
On your computer, it would be *somewhere/DIPCAT/RepoDIPCAT/reports*.
Otherwise, you will get an error message.
→ Please also see instructions in the yellow fields of the process codes.