

# URKUND ANALYSIS

The screenshot displays the URKUND analysis interface. On the left, document details are shown: Document ID BPTX\_2014\_2\_11230\_0\_416524\_0\_166544.docx (D28949722), Submitted on 2017-05-30 18:21 (+02:00), Receiver sis.fsv.cuni.cz, and Message Opětovně zaslání (konverze) Show full message. A red arrow points to the message text, which states: "6% of this approx. 39 pages long document consists of text present in 13 sources." A label "celková shoda" (overall similarity) points to this percentage.

In the center, a text box explains: "možnosti ignorovat některé ze shodných zdrojů (většinou chceme v případech shody s prací stejného studenta) => dojde k prepočtu celkové shody" (possibility to ignore some of the similar sources (usually we want in cases of similarity with the work of the same student) => there will be a recalculation of the overall similarity). A red arrow points from this text to the "Sources" pane on the right.

The "Sources" pane on the right lists 13 sources with columns for Rank and Path/File Name. A red arrow points to the first source: "11230\_BP\_MSBCK\_139526\_130153445\_Bartosikova\_Dana.pdf".

At the bottom, a comparison of two document versions is shown. The left version is highlighted in yellow and has a "79%" similarity score. The right version is in white and has a "79%" similarity score. A red arrow points from the text "shoda pouze v rámci pasáže" (similarity only within the passage) to the highlighted yellow area.

- Only academics are truly competent to judge the outcome of the URKUND analysis. The supervisor usually checks the results.
- Overall percentage of similarity does not automatically indicates whether there is some problem with the thesis
  - There can be e.g. 90% of sameness found, but no actual problem
    - (this usually concerns similarity with the same text uploaded in the past)
  - Conversely, there can be a similarity ratio of 5% and an actual problem
  - ⇒ it is not the percentage that usually matters, but whether the quoted sources has been appropriately acknowledged
- You can find a **Sources** pane on the top right hand corner:
  - a list of similar sources found appears bellow
  - **there is a checkbox at the end of each line by which you can ignore a certain source from the analysis**

- this way an overall similarity percentage of 90% can drop down all the way to e.g. 2%
- **The percentage appearing in the top left hand corner of the yellow highlighted box informs solely about similarity ratio of the given text passage and not of the text as a whole**
  - The system indicates a similar passage and then resolves the similarity ratio of text within
- In the bottom left hand narrow area a number of layers of gray and yellow bars may appear
  - these indicate the similar passages in relation to different sources
    - (the same passage can be found similar in multiple different sources)
  - You can activate each layer by clicking on it. After activating the bar a source text changes in the bottom hand pane.
  - You can read further details about the source in its header.
    - This is usually informative especially in cases of the same text uploaded in the past, since you can see the exact date of the previous upload