

# Advanced Data Modeling

Summer Semester 2009

- Exercises VII -

*To be handed in before **2009-07-07, 10:00** via e-mail to [bercovici@uni-koblenz.de](mailto:bercovici@uni-koblenz.de) And [dividino@uni-koblenz.de](mailto:dividino@uni-koblenz.de), subject line: [ADM] ...*

Please send your eclipse project and your output file.

1. Download the input files:

We will use three ebooks from Project Gutenberg for this example:

- [The Outline of Science, Vol. 1 \(of 4\) by J. Arthur Thomson](#) [1]
- [The Notebooks of Leonardo Da Vinci](#) [2]
- [Ulysses by James Joyce](#) [3]

2. Application: Inverted Index Problem

Given the code for WordCount, modify this code to produce an inverted index of its inputs. An inverted index returns a list of documents that contain each word in those documents. Thus, if the word "cat" appears in documents A and B, but not C, then the line „cat A, B“.

[1]<http://www.gutenberg.org/etext/20417>

[2]<http://www.gutenberg.org/etext/5000>

[3]<http://www.gutenberg.org/etext/4300>