

Topic: Vocabulary mining for a software product

- ▶ **Motivation:** Software products involve domain vocabulary. For instance, a human-resources management system can be expected to refer terms of the human resources domain (employee, salary, benefit, ...). Of course, a software product will also involve technical vocabulary; think of 'printer', 'listener', 'window', etc. This research effort is concerned with the semi-automatic means of extracting both kinds of vocabularies.
- ▶ **Objective:** Study the 101companies chrestomathy as the target of vocabulary extraction. Apply extraction to both filenames and file content. Leverage existing 101companies infrastructure, where possibly. Use standard techniques in this context: stemming, stop lists, and others. Visualize the extracted vocabularies in a manner support comprehension, navigation and continuous tagging.