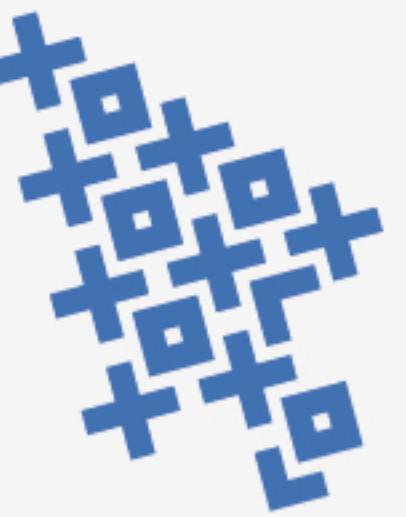


A Framework for Analysing Topics in University Courses

Ragnar Vent (Computer Science, Master's studies, Institute of Computer Science)

Supervisor: Siim Karus

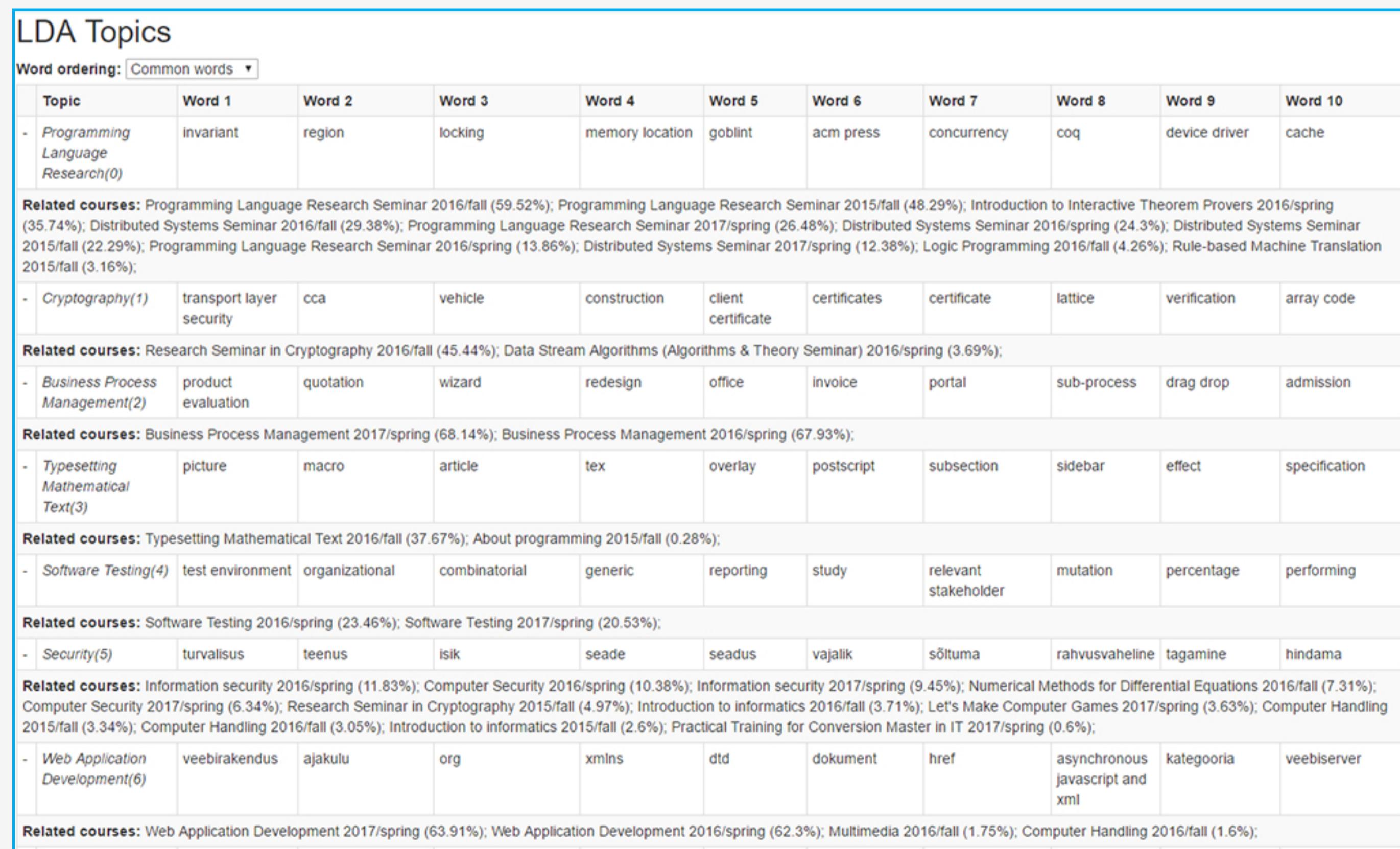


Introduction

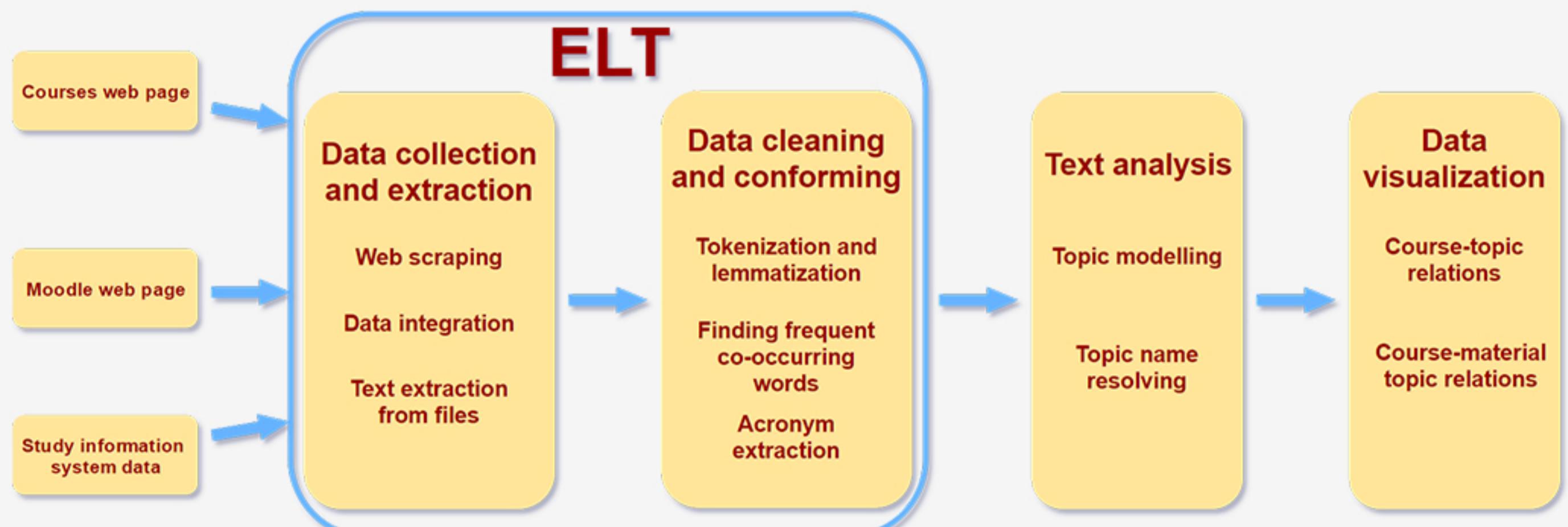
The idea behind this project was to build a framework that provides the opportunity to discover and understand topic overlaps between different courses. First, we gather course materials (e.g content pages, PDF files, PPTX files) from various sources (e.g <http://courses.cs.ut.ee/>). As a next step, all the textual data is extracted from these documents and appropriately cleansed and conformed. Finally, a topic modelling algorithm called Latent Dirichlet Allocation (LDA) is applied to the data to resolve topics.



Course topic modelling

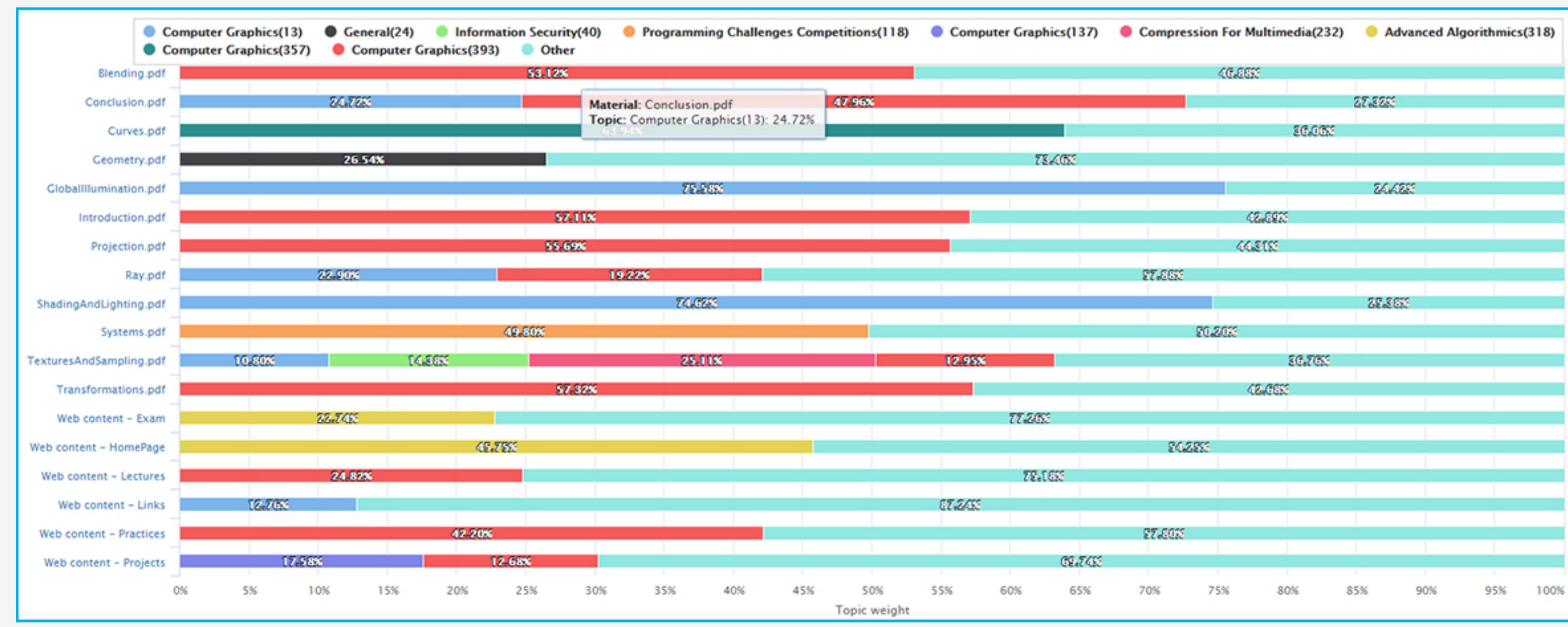


General workflow overview



atiasa.cs.ut.ee

Course material topic modelling



In addition to performing topic modelling at a course level, where the goal is to seek overlaps between different courses, we also performed topic modelling at a course material level. The aim is to investigate from which topics are individual course materials (e.g lecture slides) composed of. This opens the opportunity to seek overlaps between two course materials that are part of completely different courses.