

# Documentation

## Pre-processing

	Input	Output
<b>OutputKConversion.java</b>	author-list.txt author-key-map.txt metis.xlsx	authorkeys.txt papers.txt
Produces an assignment file where each line contains the number, name, and URI of an author. Furthermore, the Excel-file containing papers is converted to .txt and the authors' URIs in that file are converted to authors' numbers.		
<b>ProcessDBLPData.java</b>	nt-files/	dblp_papers/ dblp_confs/
Filters the information contained in the .nt-files from the DBLP-data set by eliminating data which is not associated with conferences at all. Furthermore, the remaining data is split into data referring to conference editions (dblp_confs) and data referring to publications in the context of conference editions (dblp_papers).		
<b>ProcessSemanticScholarData.java</b>	s2data/	s2_papers.txt
Reduces the amount of data retrieved from Semantic Scholar (S2) by filtering data sets with certain values for the attribute venue. The venue-names stored in S2 tend to differ from the ones stored in DBLP, thus the relevant names had to be retrieved manual and previous to running this program. Using this program, the used storage is reduced from 87.4 GB (complete S2-data sets) to 35.9 MB (potentially relevant S2-data sets).		
<b>ProcessClusters.java</b>	cluster-files/ Auth_matrix.txt $N_1 N_2 N_3 N_4$	selectedpredictions/ similarauthors.txt
Filters the clusters produced by SemEP or METIS according to the authors' numbers given as input and stores the corresponding clusters in subdirectories. Furthermore, a file is produced which contains a list of similar authors and weights for each of the authors given as input.		
<b>CollaborationFilter.java</b>	selectedclusters/ similarauthors.txt authorkeys.txt nt-files/	filteredclusters/ similarauthors_new.txt
Search the clusters in a subdirectory corresponding to a specific author for collaborations with other authors in each cluster and removes the nodes of authors who collaborated before. Thus, the result is a collaboration recommendations-network as only the association of authors remains who did not collaborate before. Moreover, authors who collaborated before are removed from the Similar-Authors-file.		

## Paper Recommendation

	Input	Output
<b>RetrieveData.java</b>	authorkeys.txt papers.txt s2_papers.txt	authors.txt papers_keywords.txt
Retrieves the keywords for each paper from the S2-data and converts the file papers.txt containing the association of Papers to Authors into a file containing the association Authors to Papers.		
<b>PaperRecommendation.java</b>	authors.txt papers_keywords.txt similarauthors.txt suffix	p_recommendations_suffix.txt p_recommendations_suffix.js
Produces a .txt-file containing paper recommendations and recommendation weights corresponding to the data given in the similar-authors-file. Furthermore, a .js-file is produced containing JSON-objects each containing the paper recommendations for an author.		

## Venue Recommendation

	Input	Output
<b>RetrieveData.java</b>	authorkeys.txt similarauthors.txt dblp_papers/ dblp_confs/	authorconfs.txt venues.txt conferences.txt
Retrieves the dblp data about conference editions and venues related to the authors contained in similarauthors.txt.		
<b>VenueRecommendation.java</b>	conferences.txt venues.txt similarauthors.txt authorkeys.txt authorconfs.txt suffix	v_recommendations_suffix.txt v_recommendations_suffix.js
Produces a .txt-file containing venue recommendations and recommendation weights corresponding to the data given in the similar-authors-file. Furthermore, a .js-file is produced containing JSON-objects each containing the venue recommendations for an author.		

## Visualization KORONA

	Input	Output
<b>ListToJS.java</b>	selectauthor.txt	selectauthor.js
Converts a text-file containing ID and names of authors into a JSON-object. In this way the information to visualize a select list is provided.		
<b>ClustersToJS.java</b>	authorkeys.txt selectedpredictions/ <i>suffix</i>	network_ <i>suffix</i> .js
The root-directory given as input contains subdirectory each corresponding to an author. Each subdirectory contains text-files which correspond to clusters in a network. This program converts the given file-structure and content into JSON-objects and stores them in a JS-file.		