

# Clustering and Creating Recommendations with Scholarly Publication Data

Mustafa Can Çavdar & Tolga Yılmaz

# Problem Description

- Scholarly websites provide list of papers
  - Google Scholar, Citeseerx, dblp
- Finding a particular paper is easy if it is known what is exactly needed.
- Related articles and related authors to a topic cannot be found easily.

# Motivation

- While researching on the web, it is wanted to find the people who are most related to the topic.
- People want to see who are in a certain cluster and who have similar works.
- It is not seen that people can see similar papers with stating a certain paper.
- It is a big amount of data to play with and it is a good opportunity to create an open source implementation.

# Methodology

- Collect CiteSeerx metadata.
  - Using an OAI harvester.
- Construct index over the metadata
- Recommendation
  - User defines his own interest fields.
  - User defines a list of articles based on these articles run similarity analysis to find related articles
  - User specifies an article; based on this article find similar articles.
- Extract authors
- Do similarity analysis on the authors and find similar people.

# Expected Results

- No expectation of efficiency.
- Depending on performance, data size may be reduced.
- Evaluation of effectiveness of results is possible.
- Seeing the structure of clusters is possible by visualisation.
- It will possible to answer such questions:
  - People clustered in «Information Retrieval» are also related to topic «X».

# THANK YOU FOR LISTENING!

Any Questions?