Background

DBLP is the largest open-access repository of scientific articles on computer science and provides metadata associated with publications, authors, and venues. We retrieved more than 6 million publications from DBLP and extracted pertinent metadata (e.g., abstracts, author affiliations, citations) from the publication texts to create the DBLP Discovery Dataset (D3). Now, on CS-Insights we devised a system (back- and front-end) to explore our dataset and uncover all the trends regarding computer science publications. As CS-Insights is an ongoing project we need to fix its open issues and extend its functionalities.

Goal

• Solve existing issues in CS-Insights-Roadmap

Tasks

• Work on project roadmap for CS-Insights
  o Backlog and additional features
• Propose extension for CS-Insights
  o Authors features (e.g., h-index)
PD05 Identifying Plagiarism of ChatGPT

Background

With the advent of advanced AI-powered language models like ChatGPT, the threat of machine-paraphrased plagiarism has become a serious concern. These models can generate text that is virtually indistinguishable from human writing, making it easy for individuals to commit plagiarism, but difficult for existing systems to detect. As these models become more accessible and widely adopted, the problem of plagiarism is expected to escalate, making it imperative for research institutions, publishers, and schools to have robust automated solutions in place.

Goal

• The goal of this project is to identify plagiarism of ChatGPT and other AI models in written works automatically.

Tasks

• Research and understand the inner workings of ChatGPT and other AI language models.
• Develop a method/training architecture for detecting generated and plagiarized text.
• Evaluate the performance of the tool using quantitative and qualitative assessments.