Abstract:

Collections of scholarly documents are usually not thought of as big data. However, large collections of scholarly documents often have many millions of publications, authors, citations, equations, figures, etc., and large scale related data and structures such as social networks, slides, data sets, etc. This leads to scholarly big data challenges, insights, methodologies, and applications.

We illustrate scholarly big data issues with examples of specialized search engines and recommendation systems that use information extraction and data mining in various areas such as computer science, chemistry, archaeology, acknowledgements, reference recommendation, collaboration recommendation, and others.