

12.1 Introduction The NTCIR Math Tasks are aimed at developing test collections for mathematical search in STEM (Science/Technology/Engineering/Mathematics) documents to facilitate and encourage research in mathematical information retrieval (MIR) (Liska et al. 2011) and its related fields (Guidi and Sacerdoti Coen 2016; Zanibbi and Blostein 2012). Mathematical formulae are important for the dissemination and communication of scientific information. They are not only used for numerical calculation but also for clarifying definitions or disambiguating explanations that are written in natural language. Despite the importance of math in technical documents, most contemporary information retrieval systems do not support users' access to mathematical formulae in target documents. One major obstacle to MIR research is the lack of readily available large-scale datasets with structured mathematical formulae, carefully designed tasks, and established evaluation methods. MIR involves searching for a particular mathematical concept, object, or result, often expressed using mathematical formulae, which—in their machine-readable

In a recent SIGIR Forum opinion article, Hiemstra et al. [28

] make the case for “Transitioning the Information Retrieval Literature to a Fully Open Access Model”, observing that various research communities thrive in such a setting. The ACL Anthology,

3

which

for nearly two decades has maintained an open archive of the computational linguistics and natural language processing literature published at various venues, is a particularly salient example, and serves as the main inspiration and basis for our initiative. After reviewing related endeavors from among the ACL Anthology and its offspring projects, we present a wider context of scholarly information utilities, both generic and specific to other fields. The table in Figure 1a compares a selection of popular services.

The ACL Anthology is an online platform that provides a curated collection of publications from the computational linguistics and natural language processing [

17

]. From a table-based overview, it enables easy access to publication lists by venue, year, or both. The ACL Anthology's open archives have enabled a thriving ecosystem of research projects on academic literature search and exploration, among them the ACL Anthology Searchbench