Link-Based Similarity Measures in Scientific Literature Data: Methods, Performance, and Applications

Sang-Wook Kim

Abstract

As the number of people who use scientific literature databases such as google scholar, citeSeer, and Microsoft academic search grows, the demand for literature retrieval services has been steadily increasing. One of the most popular retrieval services is to find a set of papers similar to a paper under consideration, which requires a measure that computes similarity between papers. Scientific literature databases exhibit two interesting characteristics that are different from general databases. First, the papers cited by old papers are often not included in the database due to technical and economic reasons. Second, a few papers cite recently-published papers. These two characteristics cause existing similarity measures to fail in at least one of the following cases: (1) measuring the similarity between old, but similar papers, (2) measuring the similarity between recent, but similar papers, and (3) measuring the similarity between two similar papers: one old and the other recent. In this talk, we address link-based similarity measures that take into account the characteristics of scientific literature databases. The talk consists of the following parts: (1) Previous and our methods for computing link-based similarity, (2) performance comparisons of link-based similarity measures, and (3) some other interesting applications to which link-based similarity measures can be applied.

Sang-Wook Kim
Department of Computer Science and Engineering
Hanyang University
Seoul
South Korea
e-mail: wook@agape.hanyang.ac.kr

M. Kosta and T. Sturm, eds. Proceedings of the MACIS 2013, Nanning, China, December 11-13, 2013.