

Review

On graph-based name disambiguation

Fan X., Wang J., Pu X., Zhou L., Lv B. Journal of Data and Information Quality (JDIQ) 2(2): 1-23, 2011. Type: Article

Date Reviewed: 03/19/12

Researchers often submit author names to digital library systems to search for papers. However, in the search results, papers by different authors with identical names are mixed together. While existing work on this problem often employs several attributes, such as the title, abstract, and author affiliation, Fan et al. developed GHOST, a system that employs co-authorship attributes only and groups the papers by the same authors into identical clusters.

GHOST consists of five steps: graph construction, valid path selection, similarity computation, name clustering, and user feedback. In the similarity computation step, the authors define similarity--"the confidence of two nodes corresponding to the same author"--and then utilize the computation scheme of the total resistance of a parallel circuit in Ohm's law. This idea is unique and will be helpful for researchers who work on social network analysis. In the name clustering step, the authors employ affinity propagation, which almost provides the same accuracy as classical agglomerative hierarchical clustering. Thus, they need to further improve the affinity propagation scheme. They evaluate the clustering scheme using precision, recall, and F-score; however, they also need to evaluate it with measures such as purity, inverse purity, and the B-cubed clustering measure, which are often used in clustering. In addition, this is the first work that employs user feedback in the name disambiguation problem. Their achieved runtime (0.5 to 2 seconds) to disambiguate an author name is reasonable. Also, GHOST was applied in CDBLP (<http://www.cdblp.cn>), a real database system.

The authors' finding that the affiliation information of each author is important will be helpful to people who develop disambiguation systems in search engines.

Reviewer: Kazunari Sugiyama

Review #: CR139983 (1208-0835)

Reproduction in whole or in part without permission is prohibited. Copyright 2012 ComputingReviews.com™

[Terms of Use](#) | [Privacy Policy](#)