Author Disambiguation Engine and Analysis of Russian Science in Web of Science (2007-2011)

Ilya Ponomarev, Pawel Sulima, Eva Darian, Ronan Sorensen, Brian Lawton, Ciaran Bolger, Etienne Godard, Victor Shyu, and Joshua Schnell

Thomson Reuters, 1455 Research Blvd, Rockville, MD (USA)
Email: ilya.ponomarev@thomsonreuters.com

Abstract

Scientometrics has become increasingly important in many forms of research management and policy making. Key to accurate bibliometric analysis is the ability to correctly link persons to their respective set of publications, balancing an optimal accuracy with speed when querying large datasets in which their publications reside.

We developed a disambiguation engine which allows us to effectively create a People Catalog of authors and their papers from large sets of Web of Science publication records. The algorithm takes into account factors such as similarities in co-authorships, emails, affiliations, names commonality, journal titles, article titles, publication dates, citations and self-citations.

We will report on the application of the disambiguation algorithm to a large corpus of publications with at least one Russian address in the 2007-2011 years. We processed 535K publication-author links from 162K papers, resulting in 127.5K distinct disambiguated authors. Statistical analysis of authors' publication sets for productivity, citations, h-indices, and collaborations on individual, organizational, and geographical levels   provides deeper insight into science management that was not possible without effective author disambiguation.