



# A Heuristic Approach to Solve Author Name Ambiguity Using Minimum Bibliographic Evidences

Sovan Bhattacharya<sup>1,2</sup> · Prasenjit Choudhury<sup>1</sup> · Subrata Nandi<sup>1</sup> · Dinesh K. Pradhan<sup>2</sup>

Received: 22 January 2023 / Accepted: 21 July 2023  
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2023

## Abstract

This article proposed a method to solve the author's name ambiguity problem using minimum available bibliographic evidence. Existing models are unable to solve many of the cases due to the unavailability of required evidence and features for resolving the conflict. Most of the works available in the literature mitigate the issue using the features such as author addresses, email-id, homepage, co-authors, etc. However, considering co-author as a feature still may have ambiguity as the co-author itself is an author. The proposed work attempts to resolve the issue with minimum available bibliographic information like the author's affiliation and publication year. A two-level heuristic method is proposed in this paper with the aforesaid minimum available features. The readily available disambiguate details of 100 authors from the ArnetMiner data-set are used to set the threshold of this proposed heuristic. The experimental analysis of proposed heuristics is performed on 20 authors of publicly available Microsoft Academic Graph (MAG) data-set. The result of this proposed heuristic outperforms when compared with other baseline approaches.

**Keywords** Author name ambiguity · Heuristics technique · Citation analysis · Information management · Text analytics · Bibliometric analysis

---

Prasenjit Choudhury, Subrata Nandi and Dinesh K. Pradhan contributed equally to this work.

<https://aminer.org/disambiguation>.  
<https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/>.

---

This article is part of the topical collection "Research Trends in Computational Intelligence" guest edited by Anshul Verma, Pradeepika Verma, Vivek Kumar Singh and S. Karthikeyan.

---

✉ Sovan Bhattacharya  
sovan.cse@gmail.com; sovan.bhattacharya@bcrec.ac.in

Prasenjit Choudhury  
prasenjit0007@yahoo.co.in

Subrata Nandi  
subrata.nandi@cse.nitdgp.ac.in

Dinesh K. Pradhan  
pdkrin@yahoo.co.in

<sup>1</sup> Department of CSE, NIT Durgapur, Mahatma Gandhi Road, Durgapur, West Bengal 713209, India

<sup>2</sup> Department of CSE, Dr. B. C. Roy Engineering College, Fuljhore Road, Durgapur, West Bengal 713206, India

## Introduction

An author's name ambiguity creates confusion in identifying a person by his/her name while analyzing the research performance of an author. The author's name ambiguity issue arises in the Digital Libraries (DLs) due to two reasons: *First*, the same scholar is recorded with a different name in different documents (homonyms name ambiguity), *Secondly*, different scholars have used the same name abbreviations in the published paper (synonyms name ambiguity).

Name ambiguity affects the estimation of several author's performance metrics like publication count, citation count, and h-index, etc. It may create great differences in bibliometric analysis specifically when measurement of author impact is taken into consideration. To show the impact of name ambiguity in the bibliographic data set, we have randomly chosen 20 authors' data from MAG and their corresponding ground truth information from their personal publication information. The histogram in Fig. 1 is used to show how name ambiguity influences the bibliometric information of an author. From Fig. 1 it is observed that the difference in publication count due to name ambiguity lies in upper quartile 118 and lower quartile 7 with a median value of 32. The