

**CLUSTERING DATA BIBLIOGRAFI MENGGUNAKAN
ALGORITMA K-MEANS DAN AUTHOR MATCHING
CLASSIFIER**

**Diajukan Untuk Melengkapi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer**



Oleh :

Robby Tri Putra

09011381621099

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA**

2020/2021

ACKNOWLEDGMENT PAGE

BIBLIOGRAPHIC DATA CLUSTERING USING K-MEANS ALGORITHM AND AUTHOR MATCHING CLASSIFIER

THESIS

Submitted to complete one of the conditions
Obtain a Bachelor Tittle I

By:

Robby Tri Putra
09011381621099

Knowing,
Head of Computer Systems Departement


Dr. Ir. Sukemi, M.T

NIP. 196612032006041001

Final Project Supervisor


Firdaus, M.Kom

NIP. 197801212008121003

HALAMAN PENGESAHAN

CLUSTERING DATA BIBLIOGRAFI MENGGUNAKAN ALGORITMA K-MEANS DAN AUTHOR MATCHING CLASSIFIER

TUGAS AKHIR

Diajukan untuk melengkapi salah satu syarat
memperoleh Gelar Sarjana I

Oleh:

Robby Tri Putra
09011381621099

Mengetahui,
Ketua Jurusan Sistem Komputer



Dr. Ir. Sukemi, M.T

NIP. 196612032006041001

Pembimbing Tugas Akhir

Firdaus, M.Kom

NIP. 197801212008121003

HALAMAN PERSETUJUAN

Pada hari Rabu, 08 Desember 2021 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Robby Tri Putra

NIM : 09011381621099

Judul : Clustering Data Bibliografi menggunakan Algoritma K-Means dan Author Matching Classifier.

Tim Penguji :

1. Ketua Sidang

Ir. Bambang Tutuko, M.T.
NIP. 196001121989031002

 28/12/2021

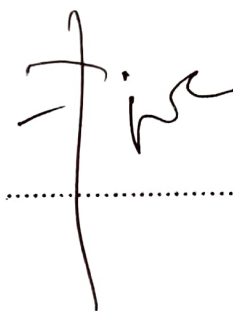
2. Penguji

Muhammad Ali Buchari, S.Kom., M.T
NIP. 198803302019031007



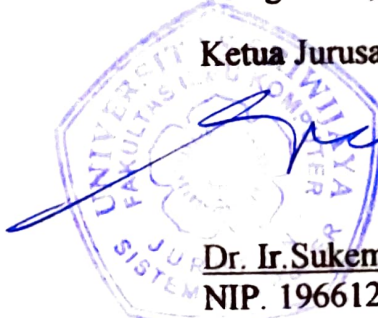
3. Pembimbing

Firdaus, M.Kom
NIP. 197801212008121003



Mengetahui,

Ketua Jurusan Sistem Komputer



Dr. Ir. Sukemi, M.T.
NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Robby Tri Putra
NIM : 09011381621099
Program Studi : Sistem Komputer
Judul Skripsi : Clustering Data Bibliografi menggunakan Algoritma K-Means dan Author Matching Classifier.

Hasil Pengecekan Software *iThenticate/Turnitin* : 15%

Menyatakan bahwa tugas akhir ini merupakan hasil karya saya sendiri dan bukan merupakan penjiplakan maupun duplikasi dari penelitian orang lain. Apabila tugas akhir ini terbukti merupakan hasil duplikasi atau penjiplakan dari hasil penelitian orang lain, maka saya bersedia menerima sanksi yang diberikan oleh akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian Surat pernyataan ini saya buat dengan sebenar-benarnya.



Palembang, Desember 2021

Menyatakan,



Robby Tri Putra

09011381621099

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji dan syukur kepada Allah SWT atas rahmat dan karunia-Nya yang besar serta shalawat dan salam penulis sampaikan kepada Nabi Muhammad SAW sehingga penulis mampu menyelesaikan penelitian dan penyusunan skripsi yang berjudul "Clustering Data Bibliografi menggunakan Algoritma K-Means dan Author Matching Classifier". Skripsi ini disusun untuk memenuhi syarat memperoleh gelar Sarjana Farmasi (S.Kom) pada Program Studi Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya.

Penelitian maupun penyusunan skripsi ini tidak lepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Pada kesempatan ini dengan segala kerendahan hati penulis mengucapkan terimakasih kepada:

1. Allah Subhanahu Wata'ala yang telah memberikan berkah serta nikmat kesehatan dan kesempatan kepada penulis dalam menyusun Proposal Tugas Akhir ini.
2. Ibu dan ayah tercinta yang selalu sedia memberikan doa, perhatian, semangat, nasihat, dan dukungan moril maupun materil kepada penulis. Saudaraku Muhammad Bayu Safikri dan Farhan Ziad Algifari yang selalu memberikan doa dan dukungan agar selalu semangat dalam mengerjakan penelitian sampai selesai.
3. Bapak Firdaus, M.Kom. selaku pembimbing Skripsi yang telah memberikan waktu, tenaga, ilmu dalam bimbingan, memberi arahan, doa, dan motivasi kepada penulis selama penelitian dan penulisan skripsi ini.
4. Bapak Deris Stiawan, M.T., Ph.D. selaku pembimbing akademik yang telah banyak meluangkan waktu, memberikan tenaga, ilmu, doa, bantuan, arahan, dan motivasi dari perkuliahan sampai tugas akhir sehingga penulis berhasil menyelesaikan skripsi ini.
5. Bapak Ir. Bambang Tutuko, M.T., Muhammad Ali Buchari, S.Kom., M.T. selaku dosen pembahas dan dosen penguji atas saran yang telah diberikan kepada penulis selama penyusunan skripsi.
6. Bapak Prof. Dr. Ir. H. Anis Saggaf, MSCE., selaku Rektor Universitas Sriwijaya, Bapak Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya. selaku Dekan Fakultas Ilmu Komputer, dan Bapak Dr. Ir. H. Sukemi, M.T.,

selaku Ketua Program Studi Sistem Komputer atas sarana dan prasarana yang telah diberikan kepada penulis sehingga penulisan skripsi ini berjalan dengan lancar.

7. Seluruh dosen Program Studi Sistem Komputer FASILKOM Universitas Sriwijaya yang telah memberikan pengetahuan, wawasan dan bantuan dalam studi penulis baik di dalam maupun di luar lingkungan kampus selama ini.
8. Mbak Renny dan mbak Sari Anhar selaku Admin Jurusan Sistem Komputer yang telah membantu dalam mengurus seluruh berkas administrasi.
9. Seluruh staf dan analis laboratorium Program Studi Sistem Komputer Universitas Sriwijaya yang telah banyak memberikan bantuan sehingga penulis dapat menyelesaikan studi dengan lancar.
10. Sahabat seperjuangan Dinar Syafina, Altundrind Wahyu, Dimas Bagus Ramarta, Ahmad Yusuf Aditama, Yoggie Al Hanif atas semua ilmu, bantuan, kebersamaan, nasihat, doa, semangat, sekaligus sebagai tempat berbagi kebahagiaan maupun kesedihan.
11. Teman-teman seperjuangan Jurusan Sistem Komputer 2016.
12. Semua pihak yang telah memberikan bantuan baik langsung maupun tidak langsung yang namanya tidak dapat disebutkan satu persatu. Penulis sangat bersyukur dan berterima kasih atas segala kebaikan, bantuan, doa, dan motivasi yang diberikan dari semua pihak yang telah membantu selamapenelitian dan penyusunan skripsi ini. Semoga Allah SWT selalu menyertakan kebaikan dan ridho-Nya kepada semua pihak yang telah membantu. Penulis menyadari dalam penulisan skripsi ini masih banyak terdapat kekurangan, oleh karena itu penulis sangat mengharapkan kritik dan saran yang membangun untuk perbaikan dimasa yang akan datang. Semoga skripsi ini dapat memberikan banyak manfaat untuk kemajuan ilmu pengetahuan.

Palembang, Desember 2021

Penulis,



Robby Tri Putra

NIM. 09011381621099

Bibliographic Data Clustering Using K-Means Algorithm and Author Matching Classifier

Robby Tri Putra (09011381621099)

Computer Engineering, Faculty of Computer, Sriwijaya University

Email : robbytriputra0@gmail.com

ABSTRAK

Author Name Disambiguation (AND) is a process to solve ambiguity problems that often occur in two or more journal publications that have author data that is almost identical or similar. The method used by the author is clustering (K-Means). K-Means is one of the methods clustering data that has a very good level of accuracy and helps solve problems in Author Name Disambiguation (AND), especially the identification of author names. The bibliographic dataset used as research material is using the DBLP.txt dataset, with this method getting a very good accuracy value of 98%.

Keywords : *Author Name Disambiguation, K-Means, Clustering, Preprocessing*

**Clustering Data Bibliografi menggunakan Algoritma K-Means dan
Author Matching Classifier**

Robby Tri Putra (09011381621099)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email : robbytriputra0@gmail.com

ABSTRAK

Author Name Disambiguation (AND) merupakan sebuah proses untuk menyelesaikan masalah ambiguitas yang sering terjadi pada dua atau lebih publikasi journal memiliki data author yang hampir mempunyai kemiripan atau sama, Metode yang digunakan penulis ialah clustering (K-Means) K-Means ialah salah satu metode clustering data yang memiliki tingkat akurasi yang sangat baik dan membantu menyelesaikan permasalahan pada Author Name Disambiguation (AND) khususnya identifikasi author name, Dataset bibliografi yang digunakan sebagai bahan penelitian adalah menggunakan Dataset DBLP.txt, dengan metode ini mendapatkan nilai akurasi yang sangat baik yaitu 98%.

Kata kunci : Author Name Disambiguation, K-Means, Clustering, Preprocessing

DAFTAR ISI

BAB I	14
PENDAHULUAN	15
1.1 Latar belakang	15
1.2 Tujuan dan Manfaat	16
1.2.1 Tujuan.....	16
1.2.2 Manfaat.....	16
1.3 Perumusan dan Batasan Masalah.....	16
1.3.1 Perumusan Masalah.....	16
1.3.2 Batasan Masalah.....	17
1.4 Metodologi Penelitian	17
1.4.1 Metode Studi Pustaka dan Literatur	17
1.4.2 Metode Konsultasi.....	17
1.4.3 Metode Pembuatan Model	17
1.4.4 Metode Pengujian dan Validasi.....	18
1.4.5 Metode Hasil dan Analisa	18
1.4.6 Metode Penarikan Kesimpulan dan Saran	18
1.5 Sistematika Penulis	18
BAB II	20
LANDASAN TEORI	20
2.1 Pendahuluan.....	20
2.2 Ambiguitas Nama	20
2.3 <i>Text Preprocessing</i>	23
2.4 <i>Jaro-Winkler Distance</i>	26
2.5 <i>Jaccard Similarity</i>	26
2.6 <i>Cosine Similarity</i>	27
2.7 <i>Machine Learning</i>	27
2.7.1 <i>Supervised Learning</i>	28
2.7.2 <i>Unsupervised Learning</i>	28
2.8 <i>Deep Learning</i>	28
2.9 Artificial Neural Network.....	29
2.9.1 Single Layer Neural Network	29
2.9.2 Multi Layer Neural Network.....	30

2.9.3	Competitive Layer Neural Network	31
2.10	<i>Min-Max Scaller</i>	31
2.11	Fungsi Aktivasi	32
2.11.1	Fungsi Aktivasi Sigmoid	32
2.11.2	Fungsi Aktivasi ReLU	33
2.11.3	Fungsi Aktivasi <i>Softmax</i>	33
2.12	Clustering Data	34
2.12.1	Algoritma <i>K-Means</i>	34
2.14	Rand Index	36
BAB III	29
METODOLOGI	37
3.1	Pendahuluan	37
3.2	Kerangka Kerja	37
3.3	Studi Literatur	38
3.4	Persiapan Dataset	39
3.5	Preprocessing Dataset	40
3.5.1	Pemilihan Atribut dan Label	41
3.5.2	<i>Case Folding</i>	42
3.5.3	<i>Punctuation Removal</i>	42
3.5.4	<i>Number Removal</i>	42
3.5.5	<i>Whitespace Removal</i>	42
3.6	Ekstraksi Fitur	42
3.7	<i>Clustering</i>	44
3.8	Evaluasi Hasil	46
BAB IV	48
HASIL DAN PEMBAHASAN	48
4.1	Pendahuluan	48
4.2	Hasil Akuisisi Data	48
4.3	Persiapan Data	48
4.4	Hasil <i>Pre-Processing</i> Dataset	49
4.5	Hasil Ekstraksi Fitur	53
4.6	<i>Clustering</i> Menggunakan <i>K-Means</i>	58

BAB V	63
KESIMPULAN	63

DAFTAR GAMBAR

Gambar 2.1 Disambiguasi <i>Author</i> Publikasi dalam 12 tahun terakhir	17
Gambar 2.2 <i>Case Folding</i>	18
Gambar 2.3 <i>Tokenizing</i>	19
Gambar 2.4 <i>Filtering</i>	19
Gambar 2.5 <i>Stemming</i>	20
Gambar 2.6 <i>Single Layer Neural Network</i>	24
Gambar 2.7 <i>Multi Layer Neural Network</i>	25
Gambar 2.8 Fungsi Aktivasi Sigmoid	26
Gambar 2.9 Fungsi Aktivasi ReLU	27
Gambar 2.10 Ilustrasi <i>Clustering K-Means</i>	30
Gambar 3.1 Kerangka Kerja Penelitian	32
Gambar 3.2 Diagram alur Preprocessing Dataset	34
Gambar 3.3 Flowchart Preprocessing.....	35
Gambar 4.4 Plot Data Berdasarkan Data Author ID	50
Gambar 4.7 Hasil Pengelompokan Data <i>K-Means</i>	51
Gambar 4.8 <i>Confusion Matrix</i>	51
Gambar 4.9. Grafik Akurasi Kluster sebanyak 20x	60

DAFTAR TABEL

Tabel 3.3. Tabel <i>Confusion Matrix</i>	40
Tabel 4.1 Dataset Yang Belum Melalui Proses <i>Pre-Processing</i>	45
Tabel 4.2. Tabel <i>Pre-Processing</i>	45
Tabel 4.3. Dataset Yang Telah Melalui Proses <i>Pre-Processing</i>	47
Tabel 4.4. Hasil Perhitungan nilai <i>predict</i> dan <i>actual</i>	50
Tabel 4.5. Hasil dari <i>Precision, Recall, F1-Score</i>	52
Tabel 4.6. Spesifikasi <i>K-Means</i> yang Digunakan	53
Tabel 4.7. Hasil dari <i>Cluster</i>	55

BAB I

PENDAHULUAN

1.1 Latar belakang

Suatu sistem yang berfungsi dengan baik untuk memperbaiki masalah disambiguasi kata dengan tepat dan akurat masih menjadi masalah inti pada penelitian *Word Sense Disambiguation* (WSD) dalam ruang lingkup *Artificial Intelligence* (AI) [1], pada WSD, suatu kata bisa memiliki arti dan makna yang berbeda karena banyaknya faktor sehingga sulit untuk menilai kata yang dimaksud ialah sama atau tidak sama baik makna serta penulisan kata, permasalahan ini juga berlaku untuk *Author Name Disambiguation* (AND) [2].

Author Name Disambiguation (AND) [2]–[4] merupakan sebuah proses untuk menyelesaikan masalah ambiguitas yang sering terjadi pada dua atau lebih publikasi *journal* memiliki data author yang hampir mempunyai kemiripan atau sama. Ambiguasi nama yang sering terjadi disebabkan karena tingkat *Synonymity* dan *Homonymity* dari nama author tersebut[5]. Pada penelitian *Author Name disambiguation* (AND) sudah banyak dilakukan oleh para peneliti [3], [6]–[8] dengan berbagai macam pendekatan atau penerapan metode serta teknik yang berbeda-beda. Penelitian dalam *Machine Learning* (ML) sudah banyak sekali dilakukan oleh para peneliti dengan metode pendekatan yang berbeda-beda dalam pengolahan data pada penelitian tentang AND.

Dalam disambiguasi nama penulis (AND), publikasi dan penulis adalah dua konsep utama. Para penulis publikasi dilambangkan dengan nama mereka, untuk itu, dalam kasus multi-penulis publikasi, daftar posisi dalam urutan tampilan di header publikasi mungkin juga menjadi relevan [9]. AND merupakan sebuah sistem pengelompokan nama author terhadap suatu publikasi penelitian dalam suatu database berdasarkan banyaknya kesalahan disambiguasi yang terjadi pada tiap data sehingga menyebabkan adanya pengambilan keputusan yang ambigu dalam menentukan author tersebut adalah orang yang sama atau tidak sama secara individual [2], [5]. Permasalahan disambiguitas nama author terutama pada data bibliography sudah menjadi masalah utama dari berbagai macam penelitian sebelumnya hingga sampai sekarang. Ambiguitas nama pada data bibliography menjadi suatu hal yang negatif yang merugikan publikasi dan mempersulit pencarian data yang dituju pada perpustakaan digital[5], [10].

Dasar penelitian tentang *Author Name Disambiguation* (AND) yang akan dilakukan adalah lanjutan serta pengembangan dari penelitian-penelitian yang sudah ada sebelumnya dengan data dan metode yang saling mendukung untuk menemukan solusi yang paling akurat dalam memecahkan masalah *Author Name Disambiguation* (AND). Salah satu metode pendekatan dalam *Machine Learning* (ML) yang relatif stabil untuk mengolah varian data dalam melakukan klasifikasi dan pengambilan keputusan adalah *K-Means Clustering*.

1.2 Tujuan dan Manfaat

1.2.1 Tujuan

Tujuan dari penelitian tugas akhir, yaitu:

1. Menyelesaikan permasalahan *Author Name Disambiguation* (AND) khususnya identifikasi Author Name dengan menggunakan *K-Means Clustering*.
2. Menentukan metode dan pendekatan yang terbaik dalam menyelesaikan masalah *Author Name Disambiguation* (AND) khususnya identifikasi Author Name.

1.2.2 Manfaat

Manfaat dari penelitian tugas akhir, yaitu membantu menyelesaikan permasalahan pada Author Name disambiguation (AND) khususnya identifikasi author name dengan menggunakan *K-Means Clustering*.

1.3 Perumusan dan Batasan Masalah

1.3.1 Perumusan Masalah

Bagaimana menentukan metode dan pendekatan yang paling akurat dalam memecahkan permasalahan pada *Author Name Disambiguation* (AND) khususnya identifikasi Author Name pada data bibliography dengan menggunakan *K-Means Clustering*.

1.3.2 Batasan Masalah

Batasan Masalah dari penelitian Tugas Akhir, yaitu:

1. Penelitian mencakup permasalahan Author Name Disambiguation (AND) khususnya identifikasi Author Name.
2. Penelitian ini menggunakan basis bahasa pemrograman yaitu bahas *Python*.
3. Dataset bibliography yang digunakan sebagai bahan penelitian adalah dataset DBLP.txt.
4. Hasil penelitian yang dilakukan dengan metodo *K-Means Clustering*.
5. Output yang dihasilkan dari penelitian ini hanya berupa nilai.

1.4 Metodologi Penelitian

Dalam penulisan Tugas Akhir ini, metodologi yang digunakan adalah sebagai berikut:

1.4.1 Metode Studi Pustaka dan Literatur

Pada metode ini, dilakukan pencarian dan pengumpulan referensi berupa literatur yang terdapat pada journal mengenai tentang *Author Name Disambiguation* (AND) dengan menggunakan *K-Means Clustering*.

1.4.2 Metode Konsultasi

Pada metode ini penulis melakukan konsultasi baik secara langsung maupun tidak langsung kepada semua pihak yang telah mempunyai pengetahuan serta wawasan yang baik dalam mengatasi permasalahan yang ditemui pada penulisan Tugas Akhir ini tentang *Author Name Disambiguasi* (AND) dengan menggunakan *K-Means Clustering*.

1.4.3 Metode Pembuatan Model

Metode selanjutnya adalah membuat suatu rancangan pemodelan dengan menggunakan bahasa pemrograman *Python*.

1.4.4 Metode Pengujian dan Validasi

Pengujian dan Validasi terhadap sistem yang telah dibuat perlu dilakukan untuk melihat batasan-batasan kinerja sistem tersebut dapat menghasilkan nilai akurasi yang akurat atau sebaliknya.

1.4.5 Metode Hasil dan Analisa

Hasil dari pengujian pada Tugas Akhir ini akan dianalisa seluruh kelebihan dan kekurangannya, sehingga dapat digunakan sebagai referensi yang baik untuk peneliti selanjutnya.

1.4.6 Metode Penarikan Kesimpulan dan Saran

Metode ini merupakan tahapan akhir dari metodologi penelitian berdasarkan hasil dan analisa penelitian yang dilakukan maka akan didapatkan kesimpulan dan saran untuk peneliti sebelumnya.

1.5 Sistematika Penulis

Dalam mempermudah menyusun tugas akhir ini dan juga membuat rangkaian isi dari setiap bab yang ada pada tugas akhir ini lebih jelas maka dibuat sistematika penulis yaitu :

BAB 1 – PENDAHULUAN

Sebagai pondasi awal penelitian, bab ini membahas tentang Latar Belakang Masalah, Tujuan dan Manfaat, Perumusan dan Batasan Masalah, Metode Penelitian, dan Sistematika Penulis.

BAB 2 – TINJAUAN PUSTAKA

Bab 2 ini merupakan penjelasan Dasar Teori, Konsep dan Dasar yang dibutuhkan untuk memecahkan suatu masalah dalam penelitian yang dilakukan.

BAB 3 – METODOLOGI

Metodologi yang digunakan membahas secara rinci tentang teknik, metode, serta alur proses yang dilakukan dalam penelitian.

BAB 4 – HASIL DAN PEMBAHASAN

Bab 4 merupakan hasil pengujian dan analisis yang diperoleh dari penelitian serta pembahasan terhadap hasil yang telah dicapai meliputi kelebihan dan kekurangan dari penelitian yang akan dilakukan.

BAB 5 – KESIMPULAN DAN SARAN

Pada bab 5 berisi kesimpulan yang bersumber dari hasil penelitian yang dilakukan beserta saran untuk penelitian selanjutnya khusus tentang Author Name Disambiguation.

Daftar Pustaka

- [1] R. Navigli, “Word sense disambiguation: A survey,” *ACM Comput. Surv.*, vol. 41, no. 2, 2009.
- [2] N. R. Smalheiser and V. I. Torvik, “Author name disambiguation,” *Annu. Rev. Inf. Sci. Technol.*, vol. 43, pp. 1–43, 2009.
- [3] M. Shoaib, A. Daud, and T. Amjad, “Author Name Disambiguation in Bibliographic Databases: A Survey,” pp. 1–24, 2020.
- [4] Z. Yamani, S. Nurmaini, and D. P. Rini, “Author Matching Classification with Anomaly Detection Approach for Bibliometric Repository Data,” *Comput. Eng. Appl. J.*, vol. 9, no. 2, pp. 79–92, 2020.
- [5] A. A. Ferreira, M. A. Gonçalves, and A. H. F. Laender, “A brief survey of automatic methods for author name disambiguation,” *SIGMOD Rec.*, vol. 41, no. 2, pp. 15–26, 2012.
- [6] C. A. D’Angelo, C. Giuffrida, and G. Abramo, “A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 62, no. 2, pp. 257–269, 2011.
- [7] Z. YAMANI, S. NURMAINI, FIRDAUS, M. Naufal R, and W. K. SARI, “Author Matching Using String Similarities and Deep Neural Networks,” vol. 172, no. Siconian 2019, pp. 474–479, 2020.
- [8] P. Gnoyke, K. Kumar, M. Kumaresh, and S. Recognition, “Author Name Disambiguation by Clustering based on Deep Learned Pairwise Similarities,” no. May, pp. 0–12, 2020.
- [9] M. C. Müller, F. Reitz, and N. Roy, “Data sets for author name disambiguation: an empirical analysis and a new resource,” *Scientometrics*, vol. 111, no. 3, pp. 1467–1500, 2017.
- [10] M. Ceci, N. Japkowicz, J. Liu, G. A. Papadopoulos, and Z. W. Raś, *Preface*, vol. 11177 LNAI. Springer International Publishing, 2018.
- [11] J. Kim and J. Kim, “The impact of imbalanced training data on machine learning for author name disambiguation,” *Scientometrics*, vol. 117, no. 1, pp. 511–526, 2018.
- [12] E. Merelli and M. Luck, “Technical Forum Group on Agents in Bioinformatics,” *Knowl. Eng. Rev.*, vol. 20, no. 2, pp. 117–125, 2004.
- [13] H. Han, H. Zha, W. Xu, and C. L. Giles, “A hierarchical naive bayes mixture model for name disambiguation in author citations,” *Proc. ACM Symp. Appl. Comput.*, vol. 2, pp. 1065–1069, 2005.
- [14] H. N. Tran, T. Huynh, and T. Do, “Author name disambiguation by using deep neural network,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect.*

Notes Bioinformatics), vol. 8397 LNAI, no. PART 1, pp. 123–132, 2014.

- [15] A. A. Ferreira, A. Veloso, M. A. Gonçalves, and A. H. F. Laender, “Effective self-training author name disambiguation in scholarly digital libraries,” *Proc. ACM Int. Conf. Digit. Libr.*, pp. 39–48, 2010.