Organized by :

Supported by :

Technical Co-sponsorship :

PROCEEDING
9th EECSI 2022

9th International Conference on Electrical Engineering, Computer Science and Informatics

October 06-07, 2022
Jakarta, Indonesia

Co-organizers :

ISBN 978-623-92135-6-5

PROCEEDINGS

9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) 2022

October 6 – 7, 2022, Jakarta – Indonesia

Editors:
Mochammad Facta, Ph.D
Mohammad Syafrullah, Ph. D
Munawar Agus Riyadi, Ph. D
Imam Much Ibnu Subroto, Ph. D
Irawan, M.Kom

PROCEEDINGS

9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) 2022

***This publication is a representation of what appears in the IEEE Digital Libraries. Some format issues inherent in the e-media version may also appear in this print version.

IEEE Catalog Number : CFP22B51-PRT, ISBN : 978-623-92135-5-8 (PRINT)
IEEE Catalog Number : CFP22B51-ART, ISBN : 978-623-92135-6-5 (DIGITAL/XPLORE COMPLIANT)

Additional Copies of This Publication Are Available From:
Curran Associates, Inc
57 Morehouse Lane Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

Foreword from General Chair EECSI 2022

In the name of Allah, Most Gracious, Most Merciful.

Welcome to the nineth International Conference on Electrical Engineering, Computer Science and Informatics (EECSI 2022) in Jakarta, Indonesia.

The 9th EECSI 2022 provides platform for researchers, academicians, professionals, and students from various engineering fields and with cross-disciplinary working or interested in the field of Electrical Engineering, Computer Science, and Informatics to share and to show their works and findings to the world.

I would like to express my hearty gratitude to all participants for coming, sharing and presenting your experiences in this vast conference. Only high-quality selected papers are accepted to be presented in this event, so we are also thankful to all the international reviewers and steering committee for their valuable work. I would like to give a compliment to all partners in publications and sponsorships for their valuable supports.

Organizing such an prestigious conference was incredibly challenging and would have been impossible without our outstanding committee, so I would like to extend my sincere appreciation to all committees and volunteers from Universitas Budi Luhur as a host and all colleagues from Universitas Islam Sultan Agung, Universitas Diponegoro, Universitas Sriwijaya, Universitas Ahmad Dahlan, Universitas Muhammadiyah Malang, Universiti Colleg TATI, Universiti Teknikal Malaysia Melaka (UTeM) and IAES Indonesia Section for providing me with much needed support, advice, and assistance on all aspects of the conference. A special thanks for IEEE Indonesia Section for the technical co-sponsorship during the conference. We do hope that this event will encourage the collaboration among us now and in the future.

We wish you all find opportunity to get rewarding technical program, intellectual inspiration, renew friendships and forge innovation, and that everyone enjoys this conference.

Mohammad Syafrullah, Ph. D
General Chair EECSI 2022

Foreword from IAES Indonesia Section

Bismillahirrohmannirrahim,
Assalamualaykum warohmatullahi wabarakatuh and Good Day,
Ladies and Gentlemen,

We would like to welcome our colleagues to attend the 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI 2022) held virtually in Jakarta on October 6th - 7th, 2022.

I hope this event will become a great event for researchers, engineers and professionals to strengthen ties and partnerships and their findings and development to the world in the field of electrical, computer, and informatics.

Institute Advanced Engineering and Science (IAES) collaborating with Universitas Budi Luhur, Universitas Diponegoro, Universitas Ahmad Dahlan, Universitas Gajah Mada, Universitas Islam Sultan Agung, Universitas Sriwijaya, Universitas Muhammadiyah Malang, and Universiti Teknologi Malaysia as several tops universities have successfully organized the conference nine times since year 2014. This achievement is due to valuable contributions also from our colleagues from Universitas Budi Luhur. I would like to put my sincere gratitude and appreciation for all partners, friends, organizing committee, reviewers, keynote speakers, and participants who have made this event as a key stage to show great progress to the world as today.

I would also like to extend my gratitude to Rector of Universitas Budi Luhur, academia and supporting staffs from Universitas Budi Luhur who become a main host and IEEE Indonesia section as a technical co-sponsor for EECSI 2022.

We wish you a happy conference and success always.

Thank you.

Assoc.Prof. Mochammad Facta, Ph.D
IAES – Indonesia Chapter

Foreword from Rector Universitas Budi Luhur

Distinguished Guests and Participants, Excellencies,
Ladies and Gentlemen

On behalf of the EECSI 2022 conference organizers, I would like to express my gratitude to all of you, who have come together here from various countries, for your cooperation which has enabled us to conduct a highly fruitful conference.

In this year's EECSI Conference which main theme was "Bridge Toward Industrial Revolution 4.0 and Its Applications on Electrical, Electronics, Computer Science and Informatics for Humanity", I expected that every participant to make contribution to this related field and promote mutual understanding among the participants through this event.

It is good for Budi Luhur University to learn about the excellent research done from different country regarding the conference topic. We also learned new ideas from each other, which we could adopt to further improve our work in this important area. I would like to pay my deep respect to all the participants for your positive participation.

We greatly appreciate the support we have from the EECSI conference organizing committee, to the Program Chairs, to the Program Committee for their extremely hard work for the details of important aspects of the conference programs and social activities. They have made this a very pleasant experience.

Finally, on behalf of the Conference Committee, I would like to express my appreciation to all the participants for taking time out of your busy duties to attend the event and to all your organizations for sending excellent participants to the event.

Assoc. Prof. Dr. Ir. Wendi Usino, M.Sc, MM
Rector Universitas Budi Luhur

Organizing Committee of EECSI 2022 Conference

Advisor
- Deni Mahdiana, Universitas Budi Luhur, Jakarta, Indonesia
- Nazori AZ, Universitas Budi Luhur, Jakarta, Indonesia
- Aghus Sofwan, Universitas Diponegoro, Semarang, Indonesia
- Zainudin Nawawi, Universitas Sriwijaya, Palembang, Indonesia
- Rahmat Budiarto, Albaha University, Baha, Saudi Arabia
- Novi Marlyana, Universitas Islam Sultan Agung, Semarang, Indonesia
- Andre Sugiyono, Universitas Islam Sultan Agung, Semarang, Indonesia
- Sunardi, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
- Wahyudi Hasbi, IEEE Indonesia Chair

General Chair
- Mohammad Syafrullah, Universitas Budi Luhur, Jakarta, Indonesia

General Co-Chair
- Mochammad Facta, Universitas Diponegoro, Semarang, Indonesia

Finance Chairs and Treasurer
- Wiwiek Fatmawati, Universitas Islam Sultan Agung, Semarang, Indonesia
- Widodo MS, Universitas Budi Luhur, Jakarta, Indonesia
- Martini, Universitas Budi Luhur, Jakarta, Indonesia
- Lina Handayani, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Program Chairs
- Imam Much Ibnu Subroto, Universitas Islam Sultan Agung, Semarang, Indonesia
- Deris Stiawan, Universitas Sriwijaya, Palembang, Indonesia

Publicity Chairs
- Indra Riyanto, Universitas Budi Luhur, Jakarta, Indonesia
- Irawan, Universitas Budi Luhur, Jakarta, Indonesia
- Hendri Irawan, Universitas Budi Luhur, Jakarta, Indonesia

Technical Program Chairs
- Munawar Agus Riyadi, Universitas Diponegoro, Semarang, Indonesia
- Mochammad Facta, Universitas Diponegoro, Semarang, Indonesia
- Tole Sutikno, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Local Arrangement, Exhibits & Registration Chairs
- Wiwin Windihastuty, Universitas Budi Luhur, Jakarta, Indonesia
- Titin Fatimah, Universitas Budi Luhur, Jakarta, Indonesia
- Samsinar, Universitas Budi Luhur, Jakarta, Indonesia
- Suwasti Broto, Universitas Budi Luhur, Jakarta, Indonesia
- Bambang Pujiyono, Universitas Budi Luhur, Jakarta, Indonesia
- Windarto, Universitas Budi Luhur, Jakarta, Indonesia
- Dolly Virgian Shaka Yudha Sakti, Universitas Budi Luhur, Jakarta, Indonesia
- Wasiran, Universitas Budi Luhur, Jakarta, Indonesia
- Anindya Putri Pradiptha, Universitas Budi Luhur, Jakarta, Indonesia

Technical Program Committee

Technical Program Committee

Sasikumar A	Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology	India
Hadhrami Ab Ghani	Universiti Malaysia Kelantan	Malaysia
Sharin Ab Ghani	Universiti Teknikal Malaysia Melaka	Malaysia
Ab Al-Hadi Ab Rahman	Universiti Teknologi Malaysia	Malaysia
Ghulam Abbas	GIK Institute of Engineering Sciences & Technology	Pakistan
Shahliza Abd Halim	University of Technology Malaysia	Malaysia
Malaoui Abdessamad	Sultan Moulay Slimane University of Beni Mellal	Morocco
Samsul Ariffin Abdul Karim	Universiti Malaysia Sabah	Malaysia
Norkhairani Abdul Rawi	Universiti Sultan Zainal Abidin	Malaysia
Mohd Azhar Abdul Razak	Universiti Teknologi Malaysia	Malaysia
Mouhamed Abdulla	Sheridan Institute of Technology	Canada
Mohammad Faiz Liew Abdullah	Universiti Tun Hussein Onn Malaysia (UTHM)	Malaysia
Azinoor Azida Abu Bakar	Universiti Teknologi MARA Johor	Malaysia
Iwan Adhicandra	University of Sydney	Australia
Mohd Ashraf Ahmad	Universiti Malaysia Pahang	Malaysia
Hani Ahmed	Universiti Malaysia Perlis	Malaysia

Ali Al Janaby	Ninevah University	Iraq
Omar Al saif	Northern Technical University	Iraq
Haider AL-Hashimi	Basra University	Iraq
Ahmed Al-Naib	Northern Technical University	Iraq
Karim Al-Saedi	Mustansiriyah University	Iraq
Jose Alanis	Universidad Tecnologica de Puebla	Mexico
Hamid Alasadi	IRAQ- BASRA	Iraq
Mudrik Alaydrus	Universitas Mercu Buana	Indonesia
Michele Albano	Aalborg University	Denmark
Mohammed Alghamdi	Al-Baha University	Saudi Arabia
Jawad Ali	University of Technology - Iraq	Iraq
Mohammed Ali	Osmania University (A I C T E)	India
Mohammad Alibakhshikenari	Universidad Carlos III de Madrid	Spain
Sinan Alkassar	Ninevah University	Iraq
Farah Alkhalid	University of Technology	Iraq
Manilal Amipara	Gujarat Technological University	India
Dian Andriana	Indonesian Institute of Sciences	Indonesia
Rakan Antar	Northern Technical University	Iraq
Farrukh Arslan	Purdue University	USA
Dominik Aufderheide	South Westphalia University of Applied Sciences Soest	Germany
Azizul Azizan	Universiti Teknologi Malaysia (UTM)	Malaysia

Eduard Babulak	Liberty University	USA
Ashish Bagwari	IEEE Member, UTU	India
Bakhyt Bakiyev	University of Birmingham	United Kingdom (Great Britain)
M'hamed Bakrim	University of Cadi Ayyad Marrakesh	Morocco
Nguyen Bao	University of Technology and Education, Ho Chi Minh City	Vietnam
Maushumi Barooah	Gauhati University	India
João Paulo Barraca	University of Aveiro	Portugal
Ruri Basuki	University of Dian Nuswantoro	Indonesia
Vikash Bhardwaj	DEWAN VS Group of Institutions	India
Aniruddha Bhattacharjya	Koneru Lakshmaiah Education Foundation	India
Yogesh Bhomia	Amity University	India
Puneeth Bhurat	Vijaya Sales Corporation	India
Sergey Biryuchinskiy	Vigitek, Inc.	USA
Ankur Bist	KIET, Ghaziabad	India
Tee Boon Tuan	Universiti Teknikal Malaysia Melaka	Malaysia
Angelo Bruno	Senior Member	Italy
Filipe Caldeira	Polytechnic Institute of Viseu	Portugal
Rodrigo Campos Bortoletto	Instituto Federal de São Paulo	Brazil
Alessandro Carrega	UNIGE	Italy

Maria Chiara Caschera	CNR	Italy
Mayank Chaturvedi	Griffith University	Australia
Datta Chavan	Bharati Vidyapeeth Deemed University College of Engineering, Pune	India
Noraini Che Pa	Universiti Putra Malaysia	Malaysia
Tai-Chen Chen	MAXEDA Technology	Taiwan
Theofilos Chrysikos	University of Patras	Greece
Young Mo Chung	Hansung University	Korea (South)
Jose Cordeiro	School of Technology of Setubal / I. P. S.	Portugal
Pablo Corral	Universidad Miguel Hernandez de Elche	Spain
Paolo Crippa	Marche Polytechnic University	Italy
Siriporn Dachasilaruk	Naresuan University	Thailand
Ahmad Nazri Dagang	Universiti Malaysia Terengganu	Malaysia
Narottam Das	CQUniversity Australia	Australia
Giuseppe De Francesco	Global Shares	Ireland
Sorin Ioan Deaconu	Politechnica University Timisoara	Romania
Jayanta Debnath	Marshall University	USA
George Dekoulis	Aerospace Engineering Institute	Cyprus
Tresna Dewi	Politeknik Negeri Sriwijaya	Indonesia
Giuseppe Di Lucca	University of Sannio	Italy
Luca Di Nunzio	University of Rome "Tor Vergata"	Italy

Moussa Diaf	Université Mouloud Mammri	Algeria
Nishant Doshi	PDPU	India
Nikolaos Doukas	Hellenic Army Academy	Greece
Supriya Dubey	SRM Institute of Science and Technology, Ghaziabad	India
Noha El-Ganainy	Norwegian University for Science and Technology NTNU	Norway
Noriko Etani	Kyoto University	Japan
Mochammad Facta	Diponegoro University	Indonesia
Farikhin Farikin	Diponegoro University	Indonesia
Miguel Franklin de Castro	Federal University of Ceará	Brazil
Franco Frattolillo	University of Sannio	Italy
Dhomas Hatta Fudholi	Universitas Islam Indonesia	Indonesia
Raad Sami Fyath	Al-Jadriya	Iraq
Vicente García Díaz	University of Oviedo	Spain
Dodi Garinto	Politeknik Manufaktur Astra	Indonesia
Antonios Gasteratos	Democritus University of Thrace	Greece
Mihai Gavrilas	Technical University of Iasi	Romania
Siamack Ghadimi	Lund University LTH	Sweden
Alireza Ghasempour	University of Applied Science and Technology	USA
Nurzal Effiyana Ghazali	Universiti Teknologi Malaysia	Malaysia
Amin Gholoobi	Open University of Cyprus	Cyprus

Baby Gobin	University of Mauritius	Mauritius
Bok-Min Goi	Universiti Tunku Abdul Rahman (UTAR)	Malaysia
Diogo Gomes	Universidade de Aveiro	Portugal
Renaldi Gondosubroto	GReS Studio	Indonesia
Dadang Gunawan	Universitas Indonesia	Indonesia
Brij Gupta	Asia University	Taiwan
Rohit Gupta	IIT Delhi	India
Ali Hamad	University of Baghdad	Iraq
Hedi Hamdi	University Of Manouba	Tunisia
Seng Hansun	Universitas Multimedia Nusantara	Indonesia
Hao Hao	RMIT University	Australia
Dedid Happyanto	Politeknik Elektronika Negeri Surabaya	Indonesia
Maha Harzallah	ISITCOM	Tunisia
Zulfatman Has	University of Muhammadiyah Malang	Indonesia
Taufik Hasan	Institut Teknologi Telkom	Indonesia
Norazlan Hashim	Shah Alam, Selangor	Malaysia
Muhammad Hasibuan	University Gadjah Mada	Indonesia
Iswadi Hasyim Rosma	Universitas Riau	Indonesia
Hendry Hendry	Chaoyang University of Technology	Taiwan
Roberto Carlos Herrera Lara	National Polytechnic School	Ecuador

Muhammad Suzuri Hitam	Universiti Malaysia Terengganu	Malaysia
Kenneth Hopkinson	Air Force Institute of Technology	USA
Jia Hou	Soochow University	China
Duy Huynh	Ho Chi Minh City University of Technology (HUTECH)	Vietnam
Fakrulradzi Idris	Universiti Teknikal Malaysia Melaka	Malaysia
Noor Ifada	University of Trunojoyo Madura	Indonesia
Amil Ahmad Ilham	Hasanuddin University	Indonesia
Paulus Insap Santosa	Universitas Gadjah Mada	Indonesia
Kashif Ishaque	Karachi Institute of Economics and Technology	Pakistan
Hossein Jafari	Intelligent Fusion Technology, Inc.	USA
Ramkumar Jaganathan	Dr NGP Arts and Science College	India
Sudhanshu Jha	University of Allahabad, Prayagraj, INDIA	India
U c Jha	Uttar Pradesh Technical University	India
Jin Jin	University of Toronto	Canada
Endra Joelianto	Institut Teknologi Bandung	Indonesia
Mohd Muzafa Jumidali	Universiti Teknologi MARA Pulau Pinang	Malaysia
Mohammed Kaabar	Washington State University	USA
Yasin Kabalci	Nigde Omer Halisdemir University	Turkey
Katerina Kabassi	Ionian University	Greece
Emil Kaburuan	Mercu Buana University	Indonesia

Sandeep Kakde	Y C College of Engineering	India
Raveendranathan Kalathil Chellappan	College of Engineering Thiruvananthapuram	India
Dimitrios Kallergis	University of West Attica	Greece
S Kannadhasan	Study World College of Engineering	India
Chutisant Kerdvibulvech	National Institute of Development Administration	Thailand
Nor Hisham Khamis	Universiti Teknologi Malaysia	Malaysia
Muhammad Imran Khan	University College Cork	Ireland
H Kiwan	University of Regina	Canada
Jens Klare	Fraunhofer FHR	Germany
Mushtaque Korai	Yanbu Industrial College	Saudi Arabia
Muhamad Koyimatu	Universitas Pertamina	Indonesia
Dragana Krstić	University of Niš	Serbia
Cheruku Kumar	Amity University Rajasthan	India
Puneet Kumar	SIoT Lab Santa Clara University	USA
Sandeep Kumar	Central Research Laboratory, Bharat Electronics Ltd.	India
Samir Ladaci	National Polytechnic School of Algiers	Algeria
Weng Siew Lam	Universiti Tunku Abdul Rahman (UTAR)	Malaysia
Magfirawaty Magfirawaty	Politeknik Siber dan Sandi Negara	Indonesia
TC Manjunath	Dayananda Sagar College of Engineering, Bangalore, Karnataka	India

Sukrisno Mardiyanto	Institut Teknologi Bandung	Indonesia
Nikhil Marriwala	Kurukshetra University	India
Zahéra Mekkioui	University of tlemcen	Algeria
Arif Muntasa	Trunojoyo University	Indonesia
Petrus Mursanto	Universitas Indonesia	Indonesia
Imamul Muttakin	Universitas Sultan Ageng Tirtayasa	Indonesia
Ruzelita Ngadiran	Universiti Malaysia Perlis & Centre of Excellence Advanced Computing (ADVCOMP), UniMAP	Malaysia
Muhammad Niswar	Universitas Hasanuddin	Indonesia
Kuntoro Nugroho	National Taiwan University of Science and Technology	Taiwan
Robertus Nugroho	Soegijapranata Catholic University	Australia
Özgür Özdemir	Konya Technical University	Turkey
Rosaura Palma-Orozco	Instituto Politécnico Nacional	Mexico
Giovanni Palmerini	Sapienza Università di Roma	Italy
Hilman Pardede	National Research and Innovation Agency of Indonesia	Indonesia
Shashikant Patil	Mumbai University	India
Phani Krishna Penumarthi	Gigamon	USA
Thinagaran Perumal	University Putra Malaysia	Malaysia
Ricardus Pramunendar	Universitas Dian Nuswantoro	Indonesia
Tri Priyambodo	Universitas Gadjah Mada	Indonesia
Aris Puji Widodo	Diponegoro University	Indonesia

Era Purwanto	Electronic Engineering Polytechnic Institute Of Surabaya	Indonesia
Ali Rafiei	University of Technology Sydney	Australia
Helmy Rahadian	Universitas Dian Nuswantoro	Indonesia
Harikumar Rajaguru	Bannari Amman Institute of Technology	India
Grienggrai Rajchakit	Maejo University	Thailand
Shuvendu Rana	SRM University AP	United Kingdom (Great Britain)
Priya Ranjan	Bhuvaneshvar Institute of Technology	India
Oday Ridha	University of Baghdad	Iraq
Munawar Riyadi	Diponegoro University	Indonesia
Olympia Roeva	Institute of Biophysics and Biomedical Engineering	Bulgaria
Yatendra Sahu	Maulana Azad National Institute of Technology, Bhopal	India
Yassine Salih-Alj	Al Akhawayn University	Morocco
Sayantam Sarkar	MVJ College of Engineering	India
Iwan Setyawan	Satya Wacana Christian University	Indonesia
Nadheer Shalash	Al-Mamoon University College	Iraq
Aditi Sharma	Parul University, Vadodara	India
Fatina Shukur	University of Kufa	Iraq
Preecha Somwang	Rajamangala University of Technology Isan	Thailand
Iickho Song	Korea Advanced Institute of Science and Technology	Korea (South)
Harco Leslie Hendric Spits Warnars	Bina Nusantara University	Indonesia

Deris Stiawan	University of Sriwijaya	Indonesia
Joey Suba	University of the Assumption	Philippines
Imam Much Ibnu Subroto	Universitas Islam Sultan Agung	Indonesia
Suherman Suherman	Universitas Sumatera Utara	Indonesia
Sangheethaa Sukumaran	Kerala Technological University	India
Hung-Min Sun	National Tsing Hua University	Taiwan
Andi Sunyoto	Universitas Amikom Yogyakarta	Indonesia
Nico Surantha	Bina Nusantara University	Indonesia
TH Sutikno	Institute of Advanced Engineering and Science	Indonesia
Tole Sutikno	Universitas Ahmad Dahlan	Indonesia
Wiwin Suwarningsih	National Research and Innovation Agency (BRIN)	Indonesia
Muhammad Syafrullah	Universitas Budi Luhur	Indonesia
Robert Szabolcsi	Óbuda University	Hungary
Srinivasulu Tadisetty	Kakatiya University College of Engineering and Technology	India
George Tambouratzis	Institute for Language & Speech Processing	Greece
Alessandro Testa	Ministry of Economy and Finance	Italy
Hapnes Toba	Maranatha Christian University	Indonesia
Tien Choon Toh	Universiti Tunku Abdul Rahman	Malaysia
Ming-Fong Tsai	National United University	Taiwan
Gloria Virginia	Duta Wacana Christian University	Indonesia

Matthias Vodel	University of Applied Sciences Mittweida	Germany
Theophilus Wellem	Satya Wacana Christian University	Indonesia
Eliana Werbin	Universidad Nacional de Córdoba	Argentina
Thaweesak Yingthawornsuk	King Mongkut's University of Technology Thonburi	Thailand
Phang Yook Ngor	Universiti Teknologi MARA Kampus Bandaraya Melaka	Malaysia
Anton Yudhana	Ahmad Dahlan University	Indonesia
Intan Yulita	Universitas Padjadjaran	Indonesia
Muhammad Yusuf	University of Trunojoyo, Madura	Indonesia
Peng Zhang	Stony Brook University	USA
Zhe Zhang	Aerospace Information Research Institute, Chinese Academy of Sciences	China
Piotr Zwierzykowski	Poznan University of Technology	Poland

Other reviewers

Additional Reviewers

Hadhrami Ab Ghani	Universiti Malaysia Kelantan	Malaysia
Shahliza Abd Halim	University of Technology Malaysia	Malaysia
Samsul Ariffin Abdul Karim	Universiti Malaysia Sabah	Malaysia
Norkhairani Abdul Rawi	Universiti Sultan Zainal Abidin	Malaysia
Azinoor Azida Abu Bakar	Universiti Teknologi MARA Johor	Malaysia
Mohd Ashraf Ahmad	Universiti Malaysia Pahang	Malaysia
Ali Othman Al Janaby	Ninevah University	Iraq
Karim Hashim Al-Saedi	Mustansiriyah University	Iraq
Jose David Alanis	Universidad Tecnologica de Puebla	Mexico
Hamid Alasadi	IRAQ- BASRA	Iraq
Michele Albano	Aalborg University	Denmark
Jawad K. Ali	University of Technology - Iraq	Iraq
Mohammed Mahmood Ali	Osmania University (A I C T E)	India
Sinan H. Alkassar	Ninevah University	Iraq
Farah Alkhalid	University of Technology	Iraq
Manilal Amipara	Gujarat Technological University	India
Dian Andriana	Indonesian Institute of Sciences	Indonesia
Farrukh Arslan	Purdue University	USA
Azizul Azizan	Universiti Teknologi Malaysia (UTM)	Malaysia
Eduard Babulak	Liberty University	USA
Bakhyt Bakiyev	University of Birmingham	United Kingdom (Great Britain)
M'hamed Bakrim	University of Cadi Ayyad Marrakesh	Morocco
Ruri Basuki	University of Dian Nuswantoro	Indonesia
Vikash Bhardwaj	DEWAN VS Group of Institutions	India
Yogesh Bhomia	Amity University	India
Sergey B. Biryuchinskiy	Vigitek, Inc.	USA

Ankur Bist	KIET, Ghaziabad	India
Tee Boon Tuan	Universiti Teknikal Malaysia Melaka	Malaysia
Angelo Bruno	Senior Member	Italy
Rodrigo Campos Bortoletto	Instituto Federal de São Paulo	Brazil
Maria Chiara Caschera	CNR	Italy
Theofilos Chrysikos	University of Patras	Greece
Young Mo Chung	Hansung University	Korea (South)
Ali Najdet Coran	Northern Technical University	Iraq
Paolo Crippa	Marche Polytechnic University	Italy
Siriporn Dachasilaruk	Naresuan University	Thailand
Ahmad Nazri Dagang	Universiti Malaysia Terengganu	Malaysia
Sorin Ioan Deaconu	Politechnica University Timisoara	Romania
Moussa Diaf	Université Mouloud Mammri	Algeria
Esmeralda Contessa Djamal	Universitas Jenderal Achmad Yani	Indonesia
Noriko Etani	Kyoto University	Japan
Amrul Faruq	Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia	Indonesia
Raad Sami Sami Fyath	Al-Jadriya	Iraq
Mihai Gavrilas	Technical University of Iasi	Romania
Alireza Ghasempour	University of Applied Science and Technology	USA
Baby Gobin	University of Mauritius	Mauritius
Bok-Min Goi	Universiti Tunku Abdul Rahman (UTAR)	Malaysia
Renaldi Gondosubroto	GReS Studio	Indonesia
Ali H. Hamad	University of Baghdad	Iraq
Seng Hansun	Universitas Multimedia Nusantara	Indonesia
Dedid Cahya Happyanto	Politeknik Elektronika Negeri Surabaya	Indonesia
Maha Harzallah	ISITCOM	Tunisia
Zulfatman Has	University of Muhammadiyah Malang	Indonesia
Norazlan Hashim	Shah Alam, Selangor	Malaysia
Muhammad Said Hasibuan	University Gadjah Mada	Indonesia

Iswadi Hasyim Rosma	Universitas Riau	Indonesia
Hendry Hendry	Chaoyang University of Technology	Taiwan
Kenneth Mark Hopkinson	Air Force Institute of Technology	USA
Jia Hou	Soochow University	China
Fakrulradzi Idris	Universiti Teknikal Malaysia Melaka	Malaysia
Indra Indra	Universitas Budi Luhur	Indonesia
Sudhanshu Kumar Jha	University of Allahabad, Prayagraj, INDIA	India
U c Jha	Uttar Pradesh Technical University	India
Jin Jin	University of Toronto	Canada
Katerina Kabassi	Ionian University	Greece
Emil R. Kaburuan	Mercu Buana University	Indonesia
Sandeep Kakde	Y C College of Engineering	India
Raveendranathan Kalathil Chellappan	College of Engineering Thiruvananthapuram	India
ayoub E. Kamal	Ntu	Iraq
S Kannadhasan	Study World College of Engineering	India
Chutisant Kerdvibulvech	National Institute of Development Administration	Thailand
Muhammad Imran Khan	University College Cork	Ireland
H Kiwan	University of Regina	Canada
Jens Klare	Fraunhofer FHR	Germany
Mushtaque Korai	Yanbu Industrial College	Saudi Arabia
Muhamad Koyimatu	Universitas Pertamina	Indonesia
Dragana Krstić	University of Niš	Serbia
Weng Siew Lam	Universiti Tunku Abdul Rahman (UTAR)	Malaysia
Magfirawaty Magfirawaty	Politeknik Siber dan Sandi Negara	Indonesia
Sukrisno Mardiyanto	Institut Teknologi Bandung	Indonesia
Nikhil Marriwala	Kurukshetra University	India
Arief Marwanto	Universitas Islam Sultan Agung (UNISSULA) Semarang	Indonesia
Arif Muntasa	Trunojoyo University	Indonesia
Petrus Mursanto	Universitas Indonesia	Indonesia
Imamul Muttakin	Universitas Sultan Ageng Tirtayasa	Indonesia
Ruzelita Ngadiran	Universiti Malaysia Perlis & Centre of Excellence	Malaysia

	Advanced Computing (ADVCOMP), UniMAP	
Muhammad Niswar	Universitas Hasanuddin	Indonesia
Kuntoro Adi Nugroho	National Taiwan University of Science and Technology	Taiwan
Rosaura Palma-Orozco	Instituto Politécnico Nacional	Mexico
Giovanni B Palmerini	Sapienza Università di Roma	Italy
Hilman F Pardede	National Research and Innovation Agency of Indonesia	Indonesia
Shashikant S. Patil	Mumbai University	India
Teguh Prakoso	Diponegoro University	Indonesia
Ricardus Anggi Pramunendar	Universitas Dian Nuswantoro	Indonesia
Era Purwanto	Electronic Engineering Polytechnic Institute Of Surabaya	Indonesia
Harikumar Rajaguru	Bannari Amman Institute of Technology	India
Grienggrai Rajchakit	Maejo University	Thailand
Priya Ranjan	Bhuvaneshvar Institute of Technology	India
Sirimonpak S	KMUTT	Thailand
Arshad Salih	Lect.	Iraq
Iwan Setyawan	Satya Wacana Christian University	Indonesia
Nadheer A. Shalash	Al-Mamoon University College	Iraq
Aditi Sharma	Parul University, Vadodara	India
Fatina Shukur	University of Kufa	Iraq
Harco Leslie Hendric Spits Warnars	Bina Nusantara University	Indonesia
Joey Suba	University of the Assumption	Philippines
Chudanat Sudthongkhong	KMUTT	Thailand
Suherman Suherman	Universitas Sumatera Utara	Indonesia
Hung-Min Sun	National Tsing Hua University	Taiwan
Andi Sunyoto	Universitas Amikom Yogyakarta	Indonesia
Nico Surantha	Bina Nusantara University	Indonesia
Wiwin Suwarningsih	National Research and Innovation Agency (BRIN)	Indonesia
George Tambouratzis	Institute for Language & Speech Processing	Greece
Alessandro Testa	Ministry of Economy and Finance	Italy
Hapnes Toba	Maranatha Christian University	Indonesia
Triwiyanto Triwiyanto	Politeknik Kesehatan Surabaya, Kementerian	Indonesia

	Kesehatan RI	
Gloria Virginia	Duta Wacana Christian University	Indonesia
Theophilus Wellem	Satya Wacana Christian University	Indonesia
Eliana Werbin	Universidad Nacional de Córdoba	Argentina
Thaweesak Yingthawornsuk	King Mongkut's University of Technology Thonburi	Thailand
Phang Yook Ngor	Universiti Teknologi MARA Kampus Bandaraya Melaka	Malaysia
Zhe Zhang	Aerospace Information Research Institute, Chinese Academy of Sciences	China

Papers by title

A

A dynamic task scheduling model for mobile cloud computing

A Novel Conformal MIMO Antenna Array based a Cylindrical Configuration for 5G Applications

An Information Security Policy Development Process in Higher Education Institution: A Case Study

An integrated model based on the location-based services for Higher Educational Systems LBS-HES

An Investigation of Ultrashort Pulses Propagation in Atmospheric Turbulence of FSO Communications

Analysis of Job Placement Based on Employee Competency Using Profile Matching

Analysis of Starting Current And Electrical Energy In Three Phase Induction Motor As A Chemical Processing System In. PT RIAU ANDALAN PULP & PAPER

Analysis Perceptions Regarding Student Exchange Using Simple Random Sampling and AHP Methods

Anomaly detection on MNIST stroke simulation dataset

Author Classification on Bibliographic Data Using Capsule Networks Architecture

Auto-Generating Business Process Model From Heterogeneous Documents: A Comprehensive Literature Survey

B

Best Lecturer Decision Support System Using Method Analytical Hierarchy Process (AHP) and Method SAW

Brain Tumor Detection on Magnetic Resonance Imaging (MRI) Images Using Convolutional Neural Network

C

Carbon Monoxide Monitoring System based on IoT with Low Power Sensor Node for Indoor Applications

Classification of Covid-19 Variants Using Boosting Algorithm

Combining Dynamic K-Means and Binary Search Centroid to Optimize Clustering Results on Home Industry Datasets

Compact MIMO Antenna Array for 5G Applications

Comparative Study Of Convolutional Neural Network And Haar Cascade Performance On Mask Detection Systems Using Matlab

Comparison of Anomaly Based and Signature Based Methods in Detection of Scanning Vulnerability

Comparison of Machine Learning Performance for Covid-19 X-ray Image Classification Based on Texture Features

Comparison Of Naive Bayes And Support Vector Machine For Detecting Hoax In Indonesian Tweet Case Study Of Tweet Covid-19

Construction of Slow and Fast Field Antenna for Detecting Lightning Strikes in South Sumatera

Control and Monitoring System Design Tunnel Light Using NodeMCU and Arduino Nano and networking

Cyberattack Feature Selection using Correlation-Based Feature Selection Method in an Intrusion Detection System

D

A B C D E F H I L M N O P R S T U W

Data Fusion for Predictive Maintenance of Industrial Assets Using Digital Twin

Decision Support System using Weighting Similarity Model for Constructing Ground-Truth Dataset

Deep Learning Neural Networks Diagnosis of Power Transformer through Its DGA Data

Design and Analysis of Multi-Core Fiber with SDM (Space Division Multiplex) Technology in Singapore to Surabaya via Southern Sea of Java

Design and Implementation of an Internet of Things Based Smart Energy Meter using Radio Frequency communication protocol

Detection of Fetal Cardiac Chamber Three Vessel Trachea View using Deep Learning

Development of Autonomous Control System using Self-Organizing Map and Autoregressive Self-Organizing Map

Development of Knowledge Management System with Soft System Metodhology in Aquatic Organization

E

A B C D E F H I L M N O P R S T U W

E-counseling-based Expressive Writing Therapy Platform for Overcoming Student Mental Health Problems

F

A B C D E F H I L M N O P R S T U W

H

A B C D E F H I L M N O P R S T U W

I

A B C D E F H I L M N O P R S T U W

L

A B C D E F H I L M N O P R S T U W

M

A B C D E F H I L M N O P R S T U W

N

A B C D E F H I L M N O P R S T U W

O

A B C D E F H I L M N O P R S T U W

P

A B C D E F H I L M N O P R S T U W

R

A B C D E F H I L M N O P R S T U W

S

A B C D E F H I L M N O P R S T U W

T

A B C D E F H I L M N O P R S T U W

U

A B C D E F H I L M N O P R S T U W

W

A B C D E F H I L M N O P R S T U W

TOC

2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)

Parallel Room 1

Parallel Room 2

Parallel Room 3

Parallel Room 4

Parallel Room 5

Parallel Room 6

Author index

Decision Support System using Weighting Similarity Model for Constructing Ground-Truth Dataset

M Miftakul Amin
Department of Computer Engineering Politeknik Negeri Sriwijaya & Faculty of Engineering Universitas Sriwijaya
Palembang, 30139, Indonesia
miftakul_a@polsri.ac.id

Deris Stiawan *
Faculty of Computer Science Universitas Sriwijaya
Palembang, 30139, Indonesia
deris@unsri.ac.id

Ermatita
Faculty of Computer Science Universitas Sriwijaya
Palembang, 30139, Indonesia
ermatita@unsri.ac.id

Imam Much Ibnu Subroto
Universitas Islam Sultan Agung,
Semarang, 50112, Indonesia
imam@unissula.ac.id

Lukman
Directorat General of Higher Education Research and Technology, Jakarta,
10270, Indonesia
lukman@kemdikbud.go.id

***Abstract*— This research aims to form a ground-truth dataset in the entity-matching process used to detect duplication of records in a bibliographic database. The contribution of this research is the obtained dataset which can be used as reference in measuring and evaluating the entity matching model implemented in bibliographic databases. This aim was achieved by developing a decision support system through experts who act as decision makers in the bibliographic databases field to construct ground-truth datasets. The model used in this decision support system weights similarity by comparing each attribute of the pairwise record in the dataset. An expert who understands all characteristics of the research database can use the graphical user interface to evaluate and determine the pairwise record that meets the conditions, such as duplication of records. This research produces a ground-truth dataset using the decision support system approach.**

Keywords*— *decision support system, ground-truth dataset, bibliographic database

I. INTRODUCTION

Scientific articles are currently packaged in digital form and managed using institutional repository software to ensure that the dissemination of scientific publications can reach a wider community. According to [1], this collection of scientific databases is termed a bibliographic database, which refers to a dataset consisting of scientific journals, proceedings, books, and other scientific articles packaged in a digital form. However, according to [2], there is duplication in research databases, implying that several scientific articles can be stored with more than one data entry. One of the contributing factors is that articles can be stored in more than one institutional repository and then indexed by the indexing software as different data sources. According to [3], another cause of duplication occurring in research databases can be the extraction of metadata from imperfect scientific articles, resulting in numerous errors in the derived information. The duplication conditions in these research databases are termed by [4] redundant papers.

The detection of record duplication in a database is highly needed at the data integration stage when the data source does not have a record identifier [5]. In addition, it is used to improve the quality and data validity in a database. According to [6], when duplication detection is applied to a single database, it is known as deduplication, and if applied to a different database, it is known as record linkage. According to [7], duplication detection compares attribute values on pairs of records in a dataset and produces a quadratic comparison with a fairly large amount of data.

Various approaches and techniques have been widely used to build a model that can detect duplication in a database. To test the various models, several evaluation measures were used, such as precision, recall, and f-measure [8]. The greater the value generated in the evaluation stage is, the more precise the developed model is. This evaluation stage can be done by referring to a ground-truth dataset as a dataset containing valid data information [9]. In the duplication detection context, the ground-truth dataset contains pairs of records with duplicate states referring to the same entity [10]. Another term for a ground-truth dataset is the gold standard, which can be used to verify the developed duplicate detection model.

A dataset which represents that record pair is duplication data in processing the entity matching is not always available. Dataset like this is known as ground-truth dataset or gold standard. Even according to Christen [10] the existence of ground-truth dataset can become the reference to confirm the status of record pair if it is duplicate or not because record pair with duplication status has been recorded in ground-truth dataset. Besides, a ground-truth dataset is also adjusted with the duplication case which would be detected until it is adjusted with the domain and problem scope that would be solved. The effort of manifesting ground-truth dataset needs quite hard effort because of the adjustment with data characteristics and expertise of an expert who has capability to determine the duplication criteria from the problem which would be solved.

This research aims to build a ground-truth dataset through the decision support system approach using a weighting similarity model. This decision support model can be used to determine whether a pair of records is duplicate. Duplication determination is recommended by the decision support system and validated by an expert who understands bibliographic database management. This ground-truth dataset can later be used as a reference dataset to measure duplication detection performance (e.g., precision, recall, and f-measure).

II. Literature Review

The use of the similarity function in research on deduplication in research databases was carried out by [11] using the graph approach applied in the OpenAire architecture. The similarity function was used to obtain the degree of similarity between the entities in a graph, such as the author and the title of the scientific publication. The similarity function was also used in research [12] to develop a dbDedup model that can detect duplication and perform compression to save storage and avoid remote database replication processes. Furthermore, another research [13] used various similarity functions to validate the results of duplication in databases and suggested that character-based similarity using Jaro–Winkler provided the best detection results. According to [10], the ground-truth dataset at the model evaluation stage in duplication detection can produce true positives, false positives, true negatives, and false negatives.

Further research on intelligent decision support systems has been carried out by [14] to provide recommendations for the selection of machine learning algorithms using case-based and rule-based reasoning approaches. This research helps non-expert users to obtain the necessary machine learning algorithm recommendations based on the particular requirements of the case to be solved. The similarity function was also used by [15] to provide product recommendations on e-commerce applications. The cosine similarity function was used to calculate the level of product similarity, and content-based filtering was applied to calculate the customer recommendation score. Through this combination of models, the system can provide recommendations for a series of products that fit customer preferences. Similar research was also performed by [16], who used the similarity function to provide product recommendations in e-commerce applications using the collaborative filtering approach. The research implementation of decision support systems for the selection of e-resources for digital scientific reference sources has been carried out by [17], who used the PROMETHE II model to provide users with recommendations on databases, journals, proceedings, books, and other electronic scientific sources.

III. Proposed Method

A. Dataset

In this research, the dataset was a Web of Science (WOS) research database, consisting of 6,190 record datasets. These data were obtained directly from the SINTA database used for indexing and scoring research activities in Indonesia.

B. Indexing Stage

In the entity matching process, the comparison stages are carried out in pairs to reduce the complexity of computing. Reference [18] emphasizes that there is a standard stage of indexing in entity matching during which a blocking process occurs. This is also supported by [10], who stated that the indexing process was carried out to reduce the level of complexity. Indexing is performed by reducing the number of matches or comparisons in pairs between entities.

Based on [6], when two datasets are compared to form a record pair comparison, datasets A and B are compared with each other, resulting in a |A| x |B| combination of record pairs, as expressed in (1). Meanwhile, if the comparison of records is carried out in a single dataset A, then the amount of record would be calculated as |A| x (|A|-1)/2 because the record compared to itself is eliminated (2).

$$|A| \times |B| \tag{1}$$

$$|A| \times (|A|-1)/2 \tag{2}$$

C. Similarity Function

The data used in the entity matching process are low in data quality, and according to [6], these data contain errors and a variety of typography errors, such as name data, affiliations, and addresses. Comparing the attribute values from two different data sources using only a combination of texts is not possible with a large data set. Generally, comparisons of attribute values produce either the same (match) or not the same information (not-match). Therefore, a more precise approach is needed to create a comparison function that could indicate the degree of similarity.

Furthermore, [19] asserts that the most frequent data types are text or string; therefore, string similarity can be one of the approaches used to compare the values of attributes. According to [20], the main purpose of this field–record pair comparison stage is to compare the attribute pairs individually and generate a comparison vector for further processing. Some techniques that can be used in string comparison include the following:

- Character-Based Similarity Metrics

One of the similarity functions included in the category of character-based similarity is the Jaro–Winkler function, which is a variant of the Jaro distance metric that measures the level of similarity between two strings and is often implemented in duplication detection [21]. The greater the value generated by the Jaro–Winkler function, the greater the degree of similarity. The value 0 means no resemblance, and the value 1 means that the two strings are absolutely equal. The Jaro similarity function uses the following formula:

$$sim_{jaro}(s1, s2) = \frac{1}{3}\left(\frac{c}{s1} + \frac{c}{s2} + \frac{c-t}{c}\right) \tag{3}$$

$$\text{Sim}_{winkler}\,(\text{s1,s2}) = \text{sim}_{jaro}\,(\text{s1,s2}) + (1.0 - \text{sim}_{jaro}\,(\text{s1,s2}))\; l_p \tag{4}$$

The variable c is the exact same number of characters; $s1$ is the length of string 1; $s2$ is the length of string 2; and t is the sum of the transpositions or the same character in the string. Variable l is the length of the common prefix at the beginning of the string with the maximum value of 4 characters, while p is the scaling factor constant, with its default value being 0.1.

- *Numeric Similarity Metrics*

In the similarity value calculation, numeric data types can also be calculated using the degree of similarity. If the numeric values $n1$ and $n2$ are exactly the same, they have a similarity value of $sim(n1, n2) = 1.0$ [10]. For the function to produce a more precise value, it can set the maximum value ($dmax$) of the existing value distribution. Formula 5 expresses the similarity function for numeric values:

$$sim_{\,num\_abs} = \begin{cases} 1.0 - \left(\frac{|n1-n2|}{d\,max}\right) if\, |n1 - n2| < dmax, \\ 0.0 \qquad\qquad\qquad\qquad else \end{cases} \tag{5}$$

D. Decision Support System

A computer-based system that can provide recommendations for a number of alternatives to assist in

decision making in structured and unstructured conditions is referred to as a decision support system [22]. The development of this decision support system requires data input to perform the reasoning process, relying on expertise and knowledge base to generate a recommendation [23].

IV. Result

This research developed a decision support system by utilizing weighting similarity using the Jaro–Winkler algorithm. This decision support system application can be used to validate whether or not a record pair is duplicated.

A. Dataset Web of Science Representation

The structure of the Web of Science dataset used in this research is reflected in Table I. Each record in the dataset is assigned an identifier (*id*), in this case *id_wos*, to perform tracking when a pair of records finishes comparing the weight of similarity.

TABLE I. Dataset Representation

Field	*Type*	*Null*	*Extra*
id_wos	int (9)	No	auto_increment
author_id	int (9)	No	
author_name	varchar (255)	No	
gswos_article_id	int (9)	Default Null	
gswos_ut	varchar (255)	Default Null	
gswos_doi	varchar (255)	Default Null	
gswos_title	varchar (255)	Default Null	
gswos_venue	varchar (255)	Default Null	
gswos_date_publish	date	Default Null	
gswos_cite	text	Default Null	
gswos_url	varchar (255)	Default Null	

B. Indexing Stage

The indexing stage is performed to reduce the number of candidate pairs which will later be used as record pairs to be compared in the next stage. From the 822 authors in the 6,190 recorded Web of Science datasets, it appears that there is a minimum of one publication and the maximum of 126 publications owned by the author (Fig. 1).

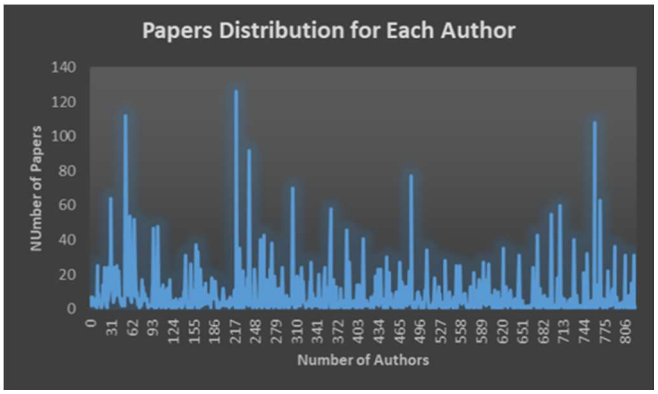


Fig. 1. Distribution of the total publication number by single author.

With the help of the record linkage toolkit library, experimental distribution of several types of blocking can be obtained, as reflected in Table II.

TABLE II. Indexing Result of Web of Science Dataset

Parameter	*Value*
Number of Record	6.190
Full Indexing	19.154.955
Sorted Neighbourhood (w=7)	21.918
Blocking (author_name)	81.117
Blocking (title)	806
Blocking (venue)	52.364

Full indexing is performed by comparing all records in the dataset using Formula 2. Therefore, the number of records is calculated as follows:

$$|A| \times (|A|-1)/2 = (6190 * 6190 - 1) / 2 = 19.154.955$$

Figure 2 summarizes the results of the indexing stage process to obtain candidate pairs. Based on the results in Table II, a total of 806 records are used as the database for the development of decision support systems.

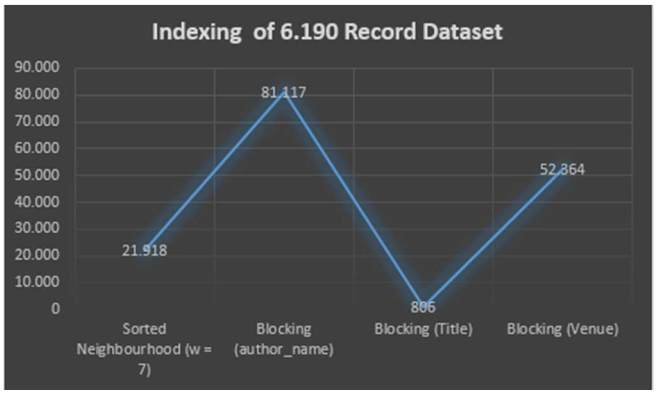


Fig. 2. The result of the indexing process based on several blocking techniques.

C. Similarity Function Implementation

After a candidate pair is obtained, the next stage is to compare the attributes that are used as determinants of whether there is a case of duplication in the research database. Four attributes are used in identification: the author, title, venue (journal/proceedings), and the period (the year of publication). Each attribute is then compared based on the level of similarity and assigned the weight of comparison. In the next stage, each weight is grouped into a weight vector (similarity sum).

The result of this similarity sum is used as the reference to determine whether there is duplication. A threshold (Θ) value will be used to determine the duplication status. If the average value of the similarity sum exceeds the threshold, it is identified as duplication (match); otherwise, if the average value of the similarity sum is below the threshold, it is not considered duplication (not match).

TABLE III. SAMPLE OF CANDIDATE PAIR SIMILARITY

id_1	id_2	author_ name	gswos_ title	gswos_ venue	gswos_date_ publish	score
6	519	0.528075	1.0	1.0	1.0	0.882019
	567	0.419192	1.0	1.0	1.0	0.854798
	2350	0.466977	1.0	1.0	1.0	0.866744
	2379	0.478355	1.0	1.0	1.0	0.869589

The score value is obtained by calculating the average value of the comparison vector as follows:

$$r6\_519 = (0.528075+1.0+1.0+1.0)/4 = 0.88201875$$
$$r6\_567 = (0.419192+1.0+1.0+1.0)/4 = 0.854798$$
$$r6\_2350 = (0.466977+1.0+1.0+1.0)/4 = 0.86674425$$
$$r6\_2379 = (0.478355+1.0+1.0+1.0)/4 = 0.86958875$$

The value of this calculation is used as the reference output for the decision support system to provide recommendations on how long this pair of records is categorized as duplication or not.

D. GUI Model of Decision Support System

An expert who understands the details of managing bibliographic databases will act as the decision maker in determining whether or not the obtained pair of records is categorized as duplicates. Figure 3 shows a graphical user interface model that can be used to help decision makers to determine whether record pairs are duplicates or not.

The similarity level of a record pair can be generated by comparing each attribute using the similarity function, and the aggregation process is performed to obtain the similarity level value from the record. Furthermore, the decision maker can validate the output of the similarity function calculation results. The *Match* button is selected when the pairwise record is a pair of duplicate records; similarly, the *Not Match* button is pressed to validate that the record pairs are not duplicates. After all alternatives have been evaluated by the decision maker, the results will then be stored as a ground-truth dataset.

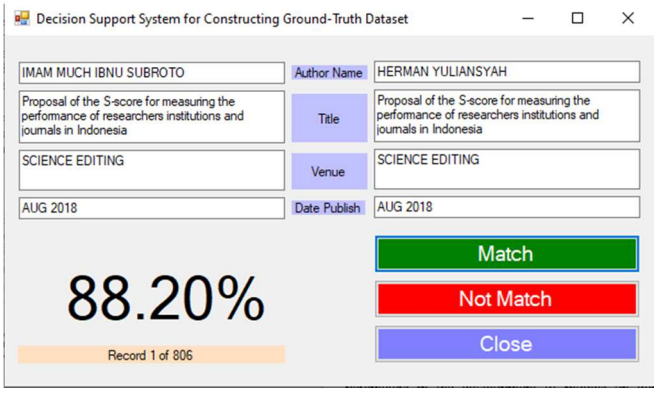


Fig. 3. The GUI of the decision support system model.

The recommendations generated by this decision support system are the first validation stage in the formation of a ground-truth dataset. The second stage of validation is performed by the decision maker as the final validator determining the record pair based on the similarity information generated by the system.

E. Co-Authorship Problem

There is information of co-authorship as the paper writing altogether then the ownership of a paper will attach to each author. However the pair of this paper author will form a unique record and not considered as duplication as seen Figure 4. Even though there is more than 1 information of paper X in the dataset is not assessed as a duplication because the information of paper X is ownership of each author.

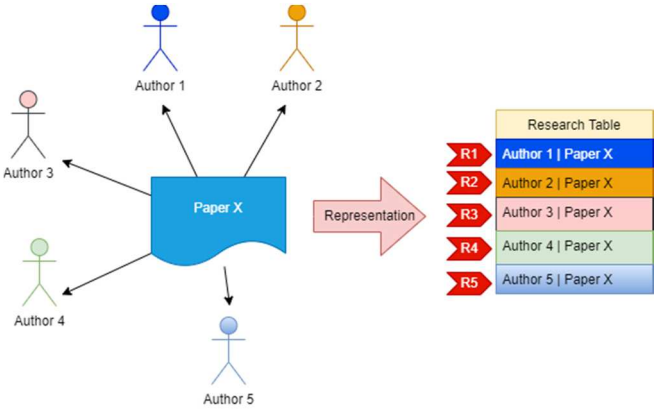


Fig. 4. Representation of Co-authorship in a Paper

Different from the condition as in Figure 5 when combination between author and paper there is more than one record. As in record 2 and record 3, where there are two same informations. Likewise in record 6, record 7, and record 8 which refer to the same information. The case like this will be identified as a duplication in the basis of research data.

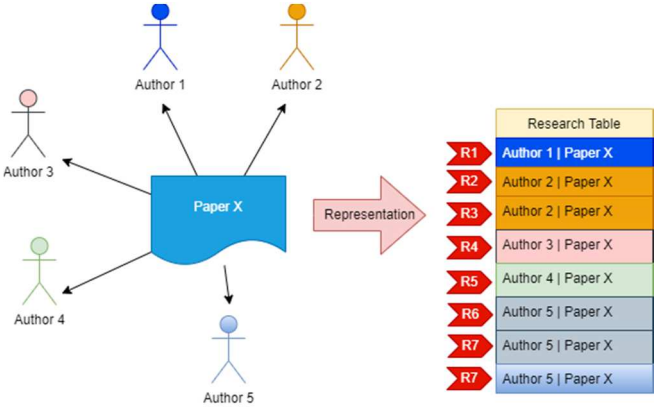


Fig. 5. Duplication Representation of co-author and paper

V. CONCLUSION AND FUTURE WORK

The proposed decision support system approach can be used to build ground-truth datasets, ensuring that the classification of bibliographic databases containing pairs of records indicated as duplicates can be determined not only by the recommendations of the decision support system but also by the experts who understand how bibliographic datasets are managed. This approach allows classification stages to be performed in layers to improve the quality of the resulting ground-truth datasets. As part of further research, scholars can utilize ground-truth datasets to measure the evaluation parameters in the development of models on duplication detection in bibliographic databases.

REFERENCES

[1] S. Kujur and A. Modak, "Study of Bibliographical Databases," *Res. Rev. Int. J. Multidiscip.*, vol. 4, no. 5, pp. 2049–2052, 2019, [Online]. Available: https://rrjournals.com/past-issue/study-of-bibliographical-databases/

[2] B. Gyawali, L. Anastasiou, and P. Knoth, "Deduplication of scholarly documents using locality sensitive hashing and word embeddings", *Lr. 2020 - 12th Int. Conf. Lang. Resour. Eval. Conf. Proc.*, no. May, pp. 901–910, 2020.

[3] A. Sefid *et al.*, "Cleaning noisy and heterogeneous metadata for record linking across scholarly big datasets", *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, pp. 9601–9606, 2019, doi: 10.1609/aaai.v33i01.33019601.

[4] X. Qi *et al.*, "Find Duplicates among the PubMed, EMBASE, and Cochrane Library Databases in Systematic Review", *PLoS One*, vol. 8, no. 8, 2013, doi: 10.1371/journal.pone.0071838.

[5] P. P. Gujar, "A survey of Record Deduplication Techniques", International Journal of Latest Trends in Engineering and Technology (IJLTET), vol. 2, no. 4, pp. 246–250, 2013.

[6] P. Christen, "A survey of indexing techniques for scalable record linkage and deduplication", *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 9, pp. 1537–1555, 2012, doi: 10.1109/TKDE.2011.127.

[7] C. Nanayakkara, P. Christen, and T. Ranbaduge, "Active Learning Based Similarity Filtering for Efficient and Effective Record Linkage", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12713 LNAI, pp. 321–333, 2021, doi: 10.1007/978-3-030-75765-6_26.

[8] C. Nanayakkara, P. Christen, T. Ranbaduge, and E. Garrett. "Evaluation measure for group-based record linkage", International Journal of Population Data Science, vol. 4, no. 1, 2019.

[9] C. Clausner, C. Papadopoulos, S. Pletschacher, and A. Antonacopoulos, "The ENP image and ground truth dataset of historical newspapers", *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2015-Novem, no. August, pp. 931–935, 2015, doi: 10.1109/ICDAR.2015.7333898.

[10] P. Christen, "*Data matching: Concepts and techniques for record linkage, entity resolution, and duplicate detection*", 2012. Springer-Verlag Berlin, doi: 10.1007/978-3-642-31164-2.

[11] P. Manghi, C. Atzori, M. De Bonis, and A. Bardi, "Entity deduplication in big data graphs for scholarly communication", *Data Technol. Appl.*, vol. 54, no. 4, pp. 409–435, 2020, doi: 10.1108/DTA-09-2019-0163.

[12] L. Xu, S. Sengupta, A. Pavlo, and G. R. Ganger, "Online deduplication for databases", *Proc. ACM SIGMOD Int. Conf. Manag. Data*, vol. Part F1277, pp. 1355–1368, 2017, doi: 10.1145/3035918.3035938.

[13] V. Wandhekar and A. Mohanpurkar, "Validation of Deduplication in Data using Similarity Measure", *Int. J. Comput. Appl.*, vol. 116, no. 21, pp. 18–22, 2015, doi: 10.5120/20460-2819.

[14] A. Mihai, "An Intelligent Decision Support System for Machine Learning Algorithms Recommendation", Master in Artificial Intelligence, Unversitat Politecnica De Catalunya, February 2017.

[15] C. Fiarni and H. Maharani, "Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods", *IJITEE (International J. Inf. Technol. Electr. Eng.*, vol. 3, no. 2, p. 42, 2019, doi: 10.22146/ijitee.45538.

[16] J. Feng, X. Fengs, N. Zhang, and J. Peng, "An improved collaborative filtering method based on similarity", *PLoS One*, vol. 13, no. 9, pp. 1–18, 2018, doi: 10.1371/journal.pone.0204003.

[17] M. Hemili, and M. R. Laouar, "E-Libraries Decision Support System for e-Resources Selection", ICIST '18: Proceedings of the 8th International Conference on Information Systems and Technologies, 2018, pp. 1-5, doi:https://doi.org/10.1145/3200842.3200857.

[18] S. Thirumuruganathan *et al.*, "Deep learning for blocking in entity matching: A design space exploration", *Proc. VLDB Endow.*, vol. 14, no. 11, pp. 2459–2472, 2021, doi: 10.14778/3476249.3476294.

[19] D. Bharambe, S. Jain, and A. Jain, "A Survey : Detection of Duplicate Record", *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 11, pp. 298–307, 2012.

[20] P. Lehti and P. Fankhauser, "Unsupervised duplicate detection using sample non-duplicates", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4244 LNCS, no. July, pp. 136–164, 2006, doi: 10.1007/11890591_5.

[21] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, "Duplicate record detection: A survey", *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 1, pp. 1–16, 2007, doi: 10.1109/TKDE.2007.250581.

[22] R. Kurniawan, A. H. Kridalaksana, and M. L. Jundillah, "Decision Support Systems Selection of Soang Superior Brood Using Weighted Product (WP) and Simple Additive Weighting (SAW) Method", E3S Web Conf. Volume 125, 2019.

[23] L. Ayu, R. Winanda, A. Arifin, F. Arrofiqi, T. W. Adi, and N. Anwar, "A design concept of fuzzy decision support system for construction workers safety monitoring", MATEC Web Conf. Volume 258, 2019.

Decision Support System using Weighting Similarity Model for Constructing Ground-Truth Dataset

By M. Miftakul Amin

Decision Support System using Weighting Similarity Model for Constructing Ground-Truth Dataset

M Miftakul Amin
Department of Computer Engineering Politeknik Negeri Sriwijaya & Faculty of Engineering Universitas Sriwijaya
Palembang, 30139, Indonesia
miftakul_a@polsri.ac.id

Deris Stiawan *
Faculty of Computer Science Universitas Sriwijaya
Palembang, 30139, Indonesia
deris@unsri.ac.id

Ermatita
Faculty of Computer Science Universitas Sriwijaya
Palembang, 30139, Indonesia
ermatita@unsri.ac.id

Imam Much Ibnu Subroto
Universitas Islam Sultan Agung,
Semarang, 50112, Indonesia
imam@unissula.ac.id

Lukman
Directorat General of Higher Education Research and Technology, Jakarta,
10270, Indonesia
lukman@kemdikbud.go.id

Abstract—This research aims to form ground-truth dataset in entity matching process which is used to detect duplication of records in bibliographic database. The benefit of this research is obtained dataset which can be used as reference in measuring and evaluating the entity matching model implemented in the bibliographic database. The approach taken is to develop a decision support system by involving experts who act as decision makers in bibliographic databases field to construct the formation of ground-truth datasets. The model used in this decision support system is weighting similarity by comparing each attribute of the pairwise record in the dataset. An expert who understands the whole characteristics of the research database can use graphical user interface to do evaluation in determining the pairwise record that meets the conditions as duplication of records. This research has produced a ground-truth dataset with decision support system approach.

Keywords—decision support system, ground-truth dataset, bibliographic database.

I. INTRODUCTION

Scientific articles are currently packaged in digital form and managed using institutional repository software, so that the dissemination of scientific publications can reach wider community. This collection of scientific databases according to [1] is called the term bibliographic database, which refers to data set consisting of scientific journals, proceedings, books, and other scientific articles packaged in digital form. However, according to [2] there is still duplication in the research database, where several scientific articles are stored with more than one number of data entries. One of the contributing factors is that an article is stored in more than one institutional repository, then indexed by indexing software as different data source. Another cause of duplication in research databases according to [3] is the extraction of metadata from imperfect scientific articles, so that the information contained in them occurs a lot of errors. The duplication conditions in these research databases, by [4] are called as redundant papers.

The detection of record duplication in the database is highly needed at the data integration stage, when the data source does not have record identifier [5]. In addition, it is also used to improve the quality and data validity in a database. According to [6] This duplication detection when applied to single database is known as deduplication, while if applied to different database it is known as record linkage. According to [7] the process that occurs in duplication detection is to compare attribute values on pairs of records in a dataset, resulting in quadratic comparison with fairly large amount of data.

Various approaches and techniques have been widely used in building a model that can detect duplication in a database. To test these various models, several evaluation measures were used such as precision, recall, and f-measure [8]. The greater the value generated in the evaluation stage, the better the model will be developed. This evaluation stage can be done by referring to a ground-truth dataset, as a dataset that contains valid data information [9]. In duplication detection context, the ground-truth dataset contains pairs of records with duplicate states, referring to the same entity [10]. Another term for this ground-truth dataset is the gold standard, which can be used to verify the developed duplicate detection model.

This research aims to build a ground-truth dataset through decision support system approach using weighting similarity model. This decision support model can be used to determine whether pair of records is duplicate or not. Duplication determination is recommended by the decision support system and validated by an expert who understands the bibliographic database management. This ground-truth dataset later can be used as a reference dataset in measuring duplication detection performance, such as to measure precision, recall, and f-measure.

II. LITERATURE REVIEW

The use of the similarity function in research on deduplication in research databases has been carried out by [11] using the graph approach applied in the OpenAire architecture. The similarity function is used to obtain the degree of similarity between entities in the graph such as the author and title of scientific publication. The similarity function was also used in research [12] to develop dbDedup model that functions to detect duplication and perform compression to save storage and remote database replication processes. Furthermore, the research performed by [13] used various similarity functions to validate the results of duplication in databases. From this research recommends that character-based similarity using jaro winkler provides the best results in detection results. Utilization of the ground-

truth dataset at the model evaluation stage in duplication detection according to [10] to produce true positives, false positives, true negatives, and false negatives.

Research on intelligent decision support systems has been done by [14] to provide recommendations for the selection of machine learning algorithms using case-based reasoning and rule-based reasoning approaches. This research has helped non-expert users to get the necessary machine learning algorithm recommendations according to the needs of the case to be solved. The use of the similarity function is also used by [15] to provide product recommendations on e-commerce applications. The cosine similarity function is used to calculate the level of product similarity, and content-based filtering to calculate the customer recommendation score. Through this combination of models, the system will provide recommendations for series of products that fit with customer preferences. Similar research was also performed by [16] using the similarity function to provide product recommendations on e-commerce applications. With collaborative filtering approach, similarity calculations are used to provide recommendations to users. Research implementation of decision support systems for the selection of e-resources for digital scientific reference sources has been carried out by [17]. The decision support system model uses PROMETHE II to provide recommendations for databases, journals, proceedings, books, and other electronic scientific sources to users.

III. PROPOSED METHOD

A. Dataset

In this research, the dataset used was web of science (WOS) research database consisting of 6,190 record datasets. This data is obtained directly from SINTA database which is database for indexing and scoring research activities in Indonesia.

B. Indexing Stage

In entity matching process, the comparison stages are carried out in pairs to reduce the complexity of computing. In this case [18] emphasizes that in entity matching there is standard stage in form of indexing in which there is blocking process. This is strength by [10] which states that to reduce this level of complexity, an indexing process is carried out. Indexing is performed by reducing the number of matches or comparisons in pairs between entities.

Technically [6] it give illustration that when two datasets are compared to form record pair comparison, then datasets A and B will be compared with each other. This will result in as much as |A| x | B| combination of record pairs as in formula (1). Meanwhile, if the comparison of records is carried out on one single dataset, |A|, then certain amount of |A| x(| A|-1)/2, because the record compared to itself will be eliminated as spelled out in formula (1).

$$|A| \times |B| \quad (1)$$

$$|A| \times (|A|-1)/2 \quad (2)$$

C. Similarity Function

The data used in the entity matching process is still low in data quality. According to [6] these data contain errors and variations of typography errors, such as name data, affiliations, addresses, and similar data. Comparing attribute values from two different data sources using only combination of text is not possible with large enough data size. In general, comparisons of attribute values produce the same information (match), and not the same (non-match). So more precise approach is needed by creating comparison function that indicates the degree of similarity.

Furthermore [19] asserts that most data is text or string data type, then string similarity is one of the approaches that can be taken to compare the values of atar attributes. According to [20] the main purpose of this field-record pair comparison stage is to compare attribute pairs individually and generate comparison vector for further processing. Some techniques that can be used in string comparison include:

- Character-Based Similarity Metrics

One of the similarity functions that include in category of character-based similarity is the jaro-winkler function which is variant of the jaro distance metric, which serves to measure the level of similarity between two strings and is often implemented in duplication detection [21]. The greater the value generated by this jaro-winkler function, the greater the degree of similarity. The value 0 means no resemblance, and the value 1 means that two strings are absolutely equal. The jaro similarity function uses formula (3).

$$sim_{jaro}(s1,s2) = \frac{1}{3}\left(\frac{c}{s1} + \frac{c}{s2} + \frac{c-t}{c}\right) \quad (3)$$

$$Sim_{winkler}(s1,s2) = sim_{jaro}(s1,s2) + (1.0 - sim_{jaro}(s1,s2))\, l_p \quad (4)$$

The variable c is the exact same number of characters, s1 is the length of the string 1, s2 is the length of the string 2, t is the sum of the transpositions or the same character on the string). The variable l is the length of the common prefix at the beginning of the string with maximum value of 4 characters, while p is scalling factor constanta with its default value being 0.1.

- *Numeric Similarity Metrics*

In similarity value calculation, numeric data types can also be calculated the degree of similarity. For numeric values n1 and n2, if they are exactly the same, they will have similarity value of sim(n1,n2) = 1.0 [10]. In order for the function to produce more precise value, it can set the maximum value (dmax) of the existing value distribution. Formula (2.5) is similarity function for numeric values.

$$sim_{num\_abs} = \begin{cases} 1.0 - \left(\frac{|n1-n2|}{d\ max}\right) if |n1-n2| < dmax, \\ 0.0 \qquad\qquad else \end{cases} \quad (5)$$

D. Decision Support System

Computer-based system that can provide recommendations for number of alternatives to assist in decision making in structured and unstructured conditions is known as decision support system [22]. The development of this decision support system requires data input to perform the reasoning process, which is based on expertise and knowledge base, where this reasoning will generate a recommendation [23].

IV. RESULT

This research developed a decision support system by utilizing weighting similarity using the jaro winkler algorithm. This decision support system application can be used to validate whether the record pair is duplication or not.

A. Dataset Web of Science Representation

The Web of Science dataset used in this research has structure as can be seen in Table I. Each record in the dataset is assigned an id as an identifier, in this case it is id_wos to do tracking when pair of records has finished comparing the weight of the similarity.

TABLE I. DATASET REPRESENTATION

Field	*Type*	*Null*	*Extra*
id_wos	int (9)	No	auto_incrment
author_id	int (9)	No	
author_name	varchar (255)	No	
gswos_article_id	int (9)	Default Null	
gswos_ut	varchar (255)	Default Null	
gswos_doi	varchar (255)	Default Null	
gswos_title	varchar (255)	Default Null	
gswos_venue	varchar (255)	Default Null	
gswos_date_publish	date	Default Null	
gswos_cite	text	Default Null	
gswos_url	varchar (255)	Default Null	

B. Indexing Stage

This indexing stage is performed to reduce the number of candidate pair which will later be used as record pairs that will be compared in the next stage. From the 822 authors contained in the 6,190 recorded web of science datasets, it appears that there is at least 1 publication owned by each author, and at most 126 publications owned by the author as can be seen in Fig. 1.

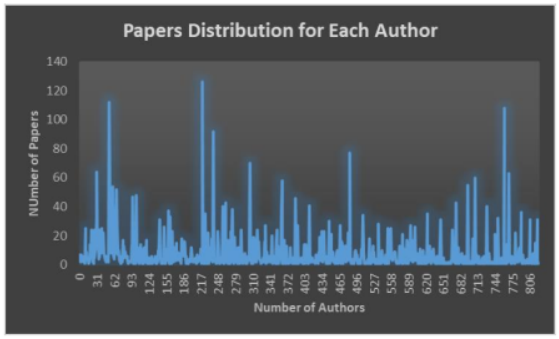


Fig. 1. Distribution of Total Paper Each Author

Using the help of record linkage toolkit library, experiment distribution of several types of blocking can be obtained as can be seen in Table II.

TABLE II. INDEXING RESULT OF WEB OF SCIENCE DATASET

Parameter	*Value*
Number of Record	6.190
Full Indexing	19.154.955
Sorted Neighbourhood (w=7)	21.918
Blocking (author_name)	81.117
Blocking (title)	806
Blocking (venue)	52.364

Full indexing is perform by comparing all records in the dataset using formula (2) so that the number of records is obtained by:

$$|A| \times (|A|-1)/2 = (6190 * 6190 - 1) / 2 = 19.154.955$$

Next fig.2 provide information on the results of indexing stage process to obtain candidate pair. So that from the results obtained in Table II, total of 806 records will be used as database for the development of decision support systems.

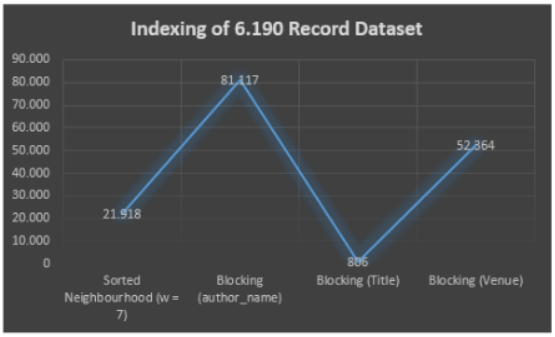


Fig. 2. Result of Indexing process from several Blocking Technique

C. Similarity Function Implementation

After the candidate pair has been obtained, the next stage is to compare the attributes that are used as determinant of whether there is case of duplication or not in the research database. there are 4 attributes used in identification, which are author, title, venue (journal / proceedings), and the period represented with information on the year of publication.

Each of these attributes will be compared to the level of similarity, and then will have weight of comparison result, and the next stage of each of these weights will be accumulated into a weight vector (similarity sum).

The result of this similarity sum is used as reference to determine whether there is duplication or not. A threshold (Θ) value will be used as threshold to determine the status of the duplication (match). If the average value of similarity sum exceeds threshold then it is identified as duplication, conversely if the average value of similarity sum is below the threshold, then it is not considered duplication (not match).

TABLE III. SAMPLE OF CANDIDATE PAIR SIMILARITY

id_1	id_2	author_ name	gswos_ title	gswos_ venue	gswos_date_ publish	score
6	519	0.528075	1.0	1.0	1.0	0.882019
	567	0.419192	1.0	1.0	1.0	0.854798
	2350	0.466977	1.0	1.0	1.0	0.866744

	2379	0.478355	1.0	1.0	1.0	0.869589

The score value is obtained from calculating the average value of the comparison vector with the following description:

r6_519 = (0.528075+1.0+1.0+1.0)/4
= 0.88201875
r6_567 = (0.419192+1.0+1.0+1.0)/4
= 0.854798
r6_2350 = (0.466977+1.0+1.0+1.0)/4
= 0.86674425
r6_2379 = (0.478355+1.0+1.0+1.0)/4
= 0.86958875

The value of this calculation will be used as reference output for the decision support system to provide recommendation on how long this pair of records is categorized as duplication or not.

D. GUI Model of Decision Support System

An expert who understands the details of managing bibliographic database will act as decision maker in determining whether the pair of records are categorized as duplicates or not. The role as decision maker will be assisted by an interface to interact with the system.

Fig. 3 shows graphical user interface model that can be used in helping decision makers to determine whether record pairs are duplicates or not. It appears that the similarity level of record pair is generated by comparing each attribute using the similarity function, and aggregation process is performed to get the similarity level value from the record. Furthermore, the decision maker can validate the output of the similarity function calculation results. The Match button is selected, when pairwise record is pair of duplicate records. While the Not Match button is used to validate that the record pairs are not duplicates. After all alternatives have been evaluated by decision maker, it will then be stored as a ground-truth dataset.

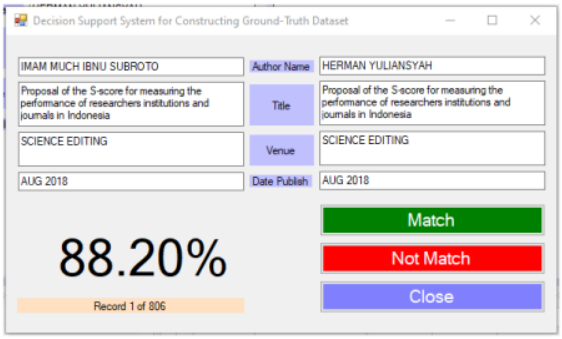


Fig. 3. GUI of decision support system model

The recommendations generated by this decision support system are the first validation in the formation of ground-truth dataset. The second stage of validation is performed by decision maker as the final validator of determining the record pair based on similarity information generated by the system.

V. Conclusion And Future Work

The decision support system approach can be used to build ground-truth dataset, so that the classification of bibliographic database containing pairs of records indicated by duplicates can be determined not only by the recommendations of the decision support system, but also by expert who understands the management of bibliographic datasets. With this approach, the classification stages are done in layers, so as to improve the quality of the resulting ground-truth dataset. As further research, researchers will utilize this ground-truth dataset to measure evaluation parameters, in the development of models on duplication detection in bibliographic databases.

References

[1] S. Kujur and A. Modak, "Study of Bibliographical Databases," *Res. Rev. Int. J. Multidiscip.*, vol. 4, no. 5, pp. 2049–2052, 2019, [Online]. Available: https://rrjournals.com/past-issue/study-of-bibliographical-databases/

[2] B. Gyawali, L. Anastasiou, and P. Knoth, "Deduplication of scholarly documents using locality sensitive hashing and word embeddings," *Lr. 2020 - 12th Int. Conf. Lang. Resour. Eval. Conf. Proc.*, no. May, pp. 901–910, 2020.

[3] A. Sefid *et al.*, "Cleaning noisy and heterogeneous metadata for record linking across scholarly big datasets," *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, pp. 9601–9606, 2019, doi: 10.1609/aaai.v33i01.33019601.

[4] X. Qi *et al.*, "Find Duplicates among the PubMed, EMBASE, and Cochrane Library Databases in Systematic Review," *PLoS One*, vol. 8, no. 8, 2013, doi: 10.1371/journal.pone.0071838.

[5] P. P. Gujar, "A survey of Record Deduplication Techniques," vol. 2, no. 4, pp. 246–250, 2013.

[6] P. Christen, "A survey of indexing techniques for scalable record linkage and deduplication," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 9, pp. 1537–1555, 2012, doi: 10.1109/TKDE.2011.127.

[7] C. Nanayakkara, P. Christen, and T. Ranbaduge, "Active Learning Based Similarity Filtering for Efficient and Effective Record Linkage," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12713 LNAI, pp. 321–333, 2021, doi: 10.1007/978-3-030-75765-6_26.

[8] J. Website, "International Journal of Evaluation measure for group-based record linkage," vol. 0, no. November, 2019.

[9] C. Clausner, C. Papadopoulos, S. Pletschacher, and A. Antonacopoulos, "The ENP image and ground truth dataset of historical newspapers," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2015-Novem, no. August, pp. 931–935, 2015, doi: 10.1109/ICDAR.2015.7333898.

[10] P. Christen, *Data matching: Concepts and techniques for record linkage, entity resolution, and duplicate detection*. 2012. doi: 10.1007/978-3-642-31164-2.

[11] P. Manghi, C. Atzori, M. De Bonis, and A. Bardi, "Entity deduplication in big data graphs for scholarly communication," *Data Technol. Appl.*, vol. 54, no. 4, pp. 409–435, 2020, doi: 10.1108/DTA-09-2019-0163.

[12] L. Xu, S. Sengupta, A. Pavlo, and G. R. Ganger, "Online deduplication for databases," *Proc. ACM SIGMOD Int. Conf. Manag. Data*, vol. Part F1277, pp. 1355–1368, 2017, doi: 10.1145/3035918.3035938.

[13] V. Wandhekar and A. Mohanpurkar, "Validation of Deduplication in Data using Similarity Measure," *Int. J. Comput. Appl.*, vol. 116, no. 21, pp. 18–22, 2015, doi: 10.5120/20460-2819.

[14] A. Mihai, "Master in Artificial Intelligence An Intelligent Decision Support System for Machine Learning Algorithms Recommendation," no. February 2017, p. 53, 2017.

[15] C. Fiarni and H. Maharani, "Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods," *IJITEE (International J. Inf. Technol. Electr. Eng.*, vol. 3, no. 2, p. 42, 2019, doi: 10.22146/ijitee.45538.

[16] J. Feng, X. Fengs, N. Zhang, and J. Peng, "An improved collaborative filtering method based on similarity," *PLoS One*, vol. 13, no. 9, pp. 1–18, 2018, doi: 10.1371/journal.pone.0204003.

[17] M. Hemili, "E-libraries decision support system for e-resources selection," pp. 1–5, 2018.

[18] S. Thirumuruganathan *et al*., "Deep learning for blocking in entity matching: A design space exploration," *Proc. VLDB Endow.*, vol. 14, no. 11, pp. 2459–2472, 2021, doi: 10.14778/3476249.3476294.

[19] D. Bharambe, S. Jain, and A. Jain, "A Survey : Detection of Duplicate Record," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 11, pp. 298–307, 2012.

[20] P. Lehti and P. Fankhauser, "Unsupervised duplicate detection using sample non-duplicates," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4244 LNCS, no. July, pp. 136–164, 2006, doi: 10.1007/11890591_5.

[21] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, "Duplicate record detection: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 1, pp. 1–16, 2007, doi: 10.1109/TKDE.2007.250581.

[22] R. Kurniawan, A. H. Kridalaksana, and M. L. Jundillah, "Decision Support Systems Selection of Soang Superior Brood Using Weighted Product (WP) and Simple Additive Weighting (SAW) Method," vol. 4, no. 201 9, 2019.

[23] L. Ayu, R. Winanda, A. Arifin, F. Arrofiqi, T. W. Adi, and N. Anwar, "A design concept of fuzzy decision support system for construction workers safety monitoring," vol. 2019, 2019.

Decision Support System using Weighting Similarity Model for Constructing Ground-Truth Dataset

ORIGINALITY REPORT

9%
SIMILARITY INDEX

PRIMARY SOURCES

1	Y A Gerhana, A R Atmadja, D S Maylawati, A Rahman, K Nufus, H Qodim, Busr, M A Ramdhani. "Computer speech recognition to text for recite Holy Quran", IOP Conference Series: Materials Science and Engineering, 2018 Crossref	27 words — 1%
2	Aziz Fellah. "All-Three: Near-optimal and domain-independent algorithms for near-duplicate detection", Array, 2021 Crossref	25 words — 1%
3	Susanto, Deris Stiawan, M. Agus Syamsul Arifin, Mohd. Yazid Idris, Rahmat Budiarto. "IoT Botnet Malware Classification Using Weka Tool and Scikit-learn Machine Learning", 2020 7th International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI), 2020 Crossref	24 words — 1%
4	hdl.handle.net Internet	23 words — 1%
5	publichealth.jmir.org Internet	21 words — 1%
6	Arief Marwanto, Yahya Hidayatullah, Imam Much Ibnu Subroto. "Experimental Analysis of IPv6	14 words — < 1%

Tunneling of Jumbo Frame Transmission using Mikrotik Routers", 2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2021
Crossref

7 Vatsalan, Dinusha, Peter Christen, and Vassilios S. Verykios. "A taxonomy of privacy-preserving record linkage techniques", Information Systems, 2012.
Crossref
14 words — < 1%

8 Rendra Gustriansyah, Ermatita, Dian Palupi Rini, Reza Firsandaya Malik. "The Criteria That Have A Significant Effect on Forecasting the Number of Sales Using the Best-Worst Method", 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), 2020
Crossref
13 words — < 1%

9 Khaerul Manaf, SW Pitara, Beki Subaeki, Rudy Gunawan, Rodiah, Bakhtiar. "Comparison of Carp Rabin Algorithm and Jaro-Winkler Distance to Determine The Equality of Sunda Languages", 2019 IEEE 13th International Conference on Telecommunication Systems, Services, and Applications (TSSA), 2019
Crossref
12 words — < 1%

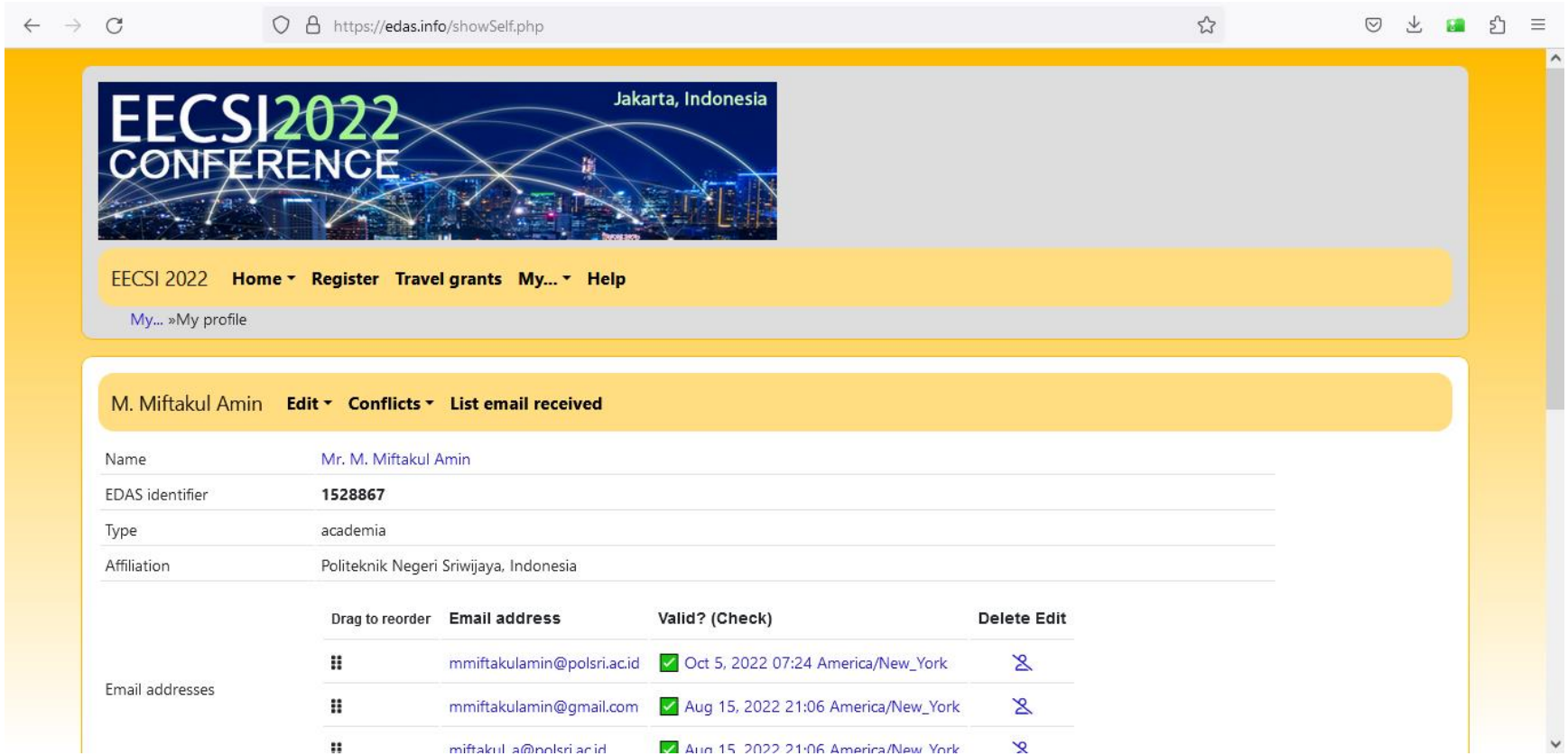
10 Proceeding of LPPM UPN “Veteran” Yogyakarta Conference Series 2020 – Engineering and Science Series, 2020
Publications
12 words — < 1%

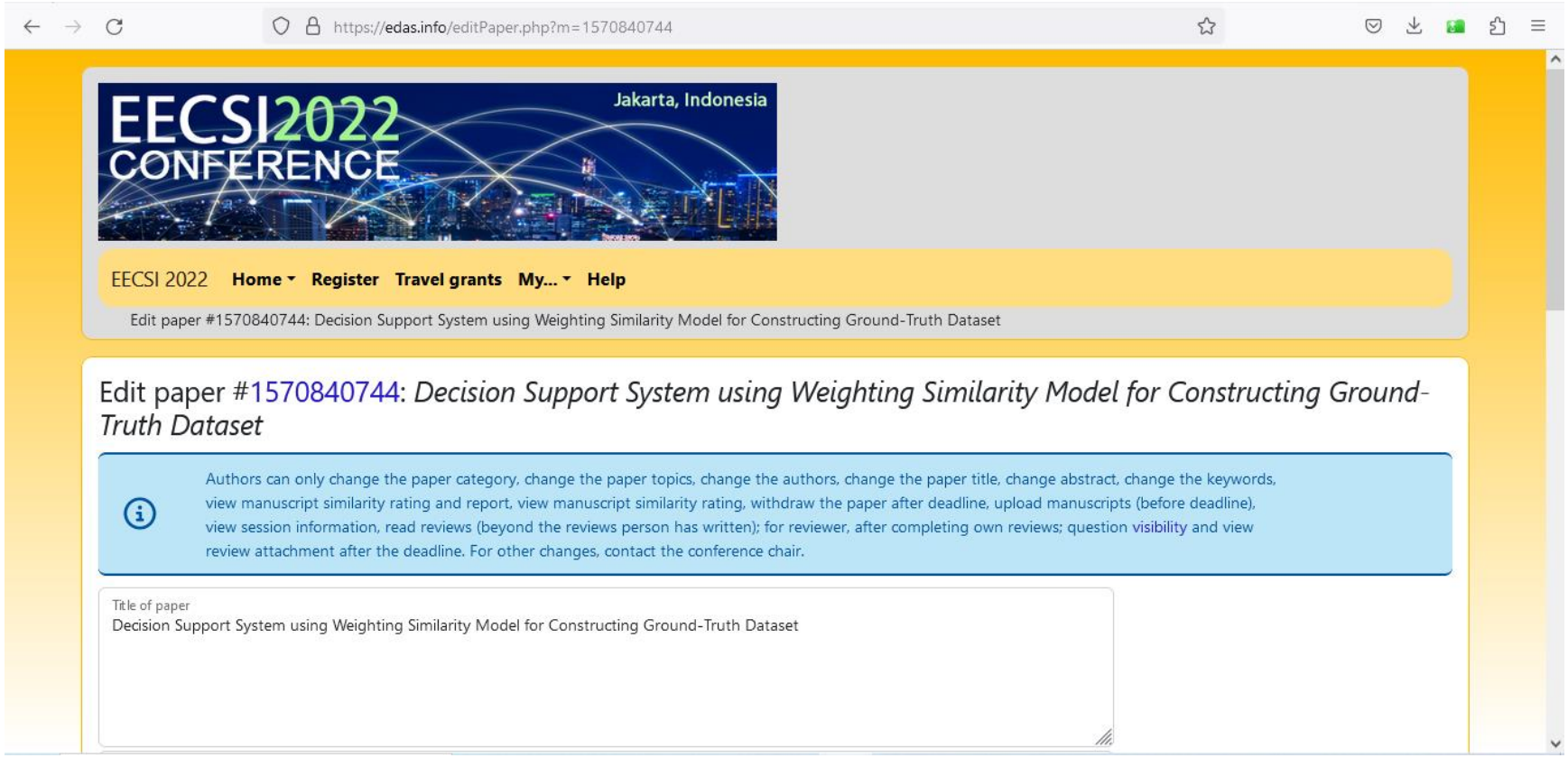
11 core.ac.uk
Internet
12 words — < 1%

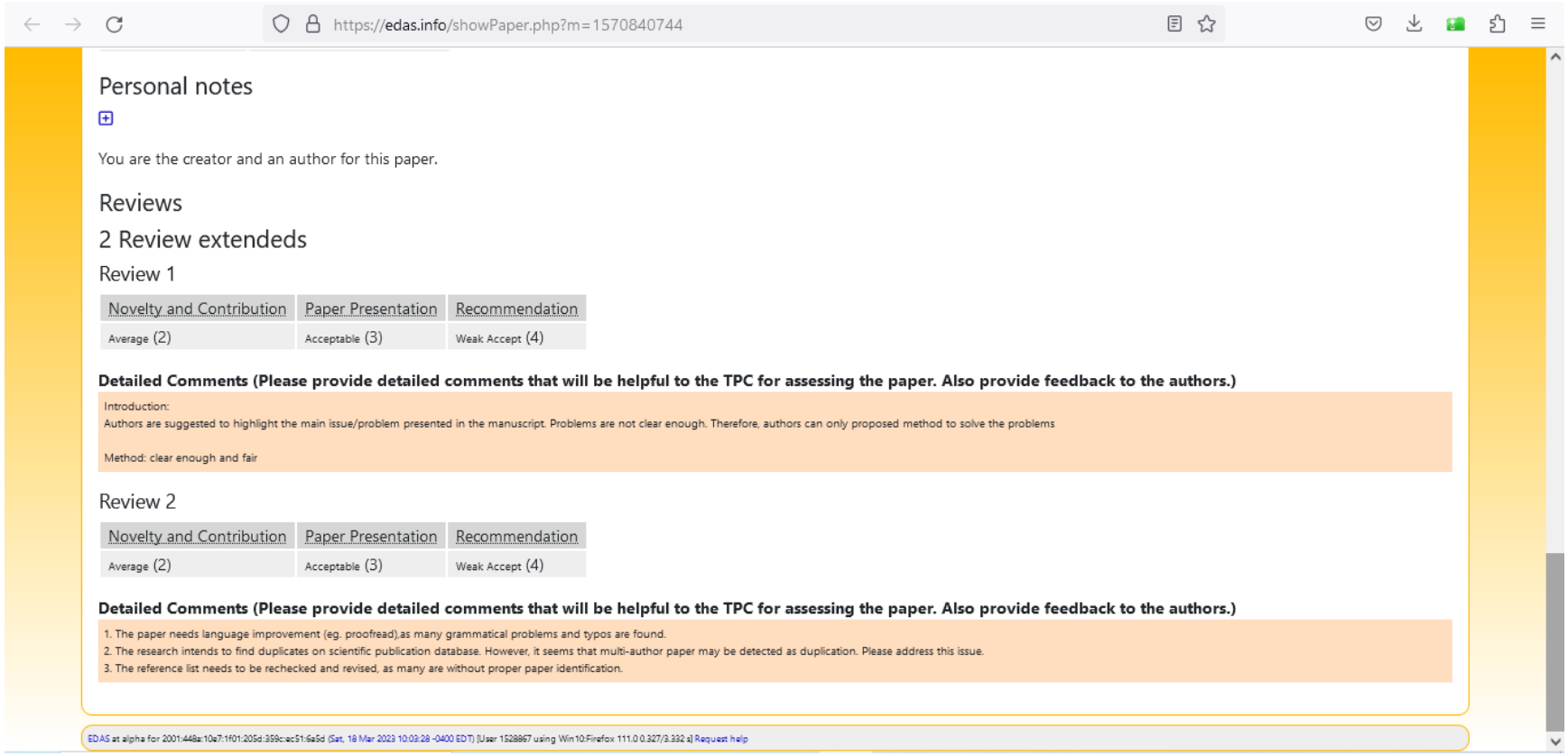
12 humaniora.journal.ugm.ac.id
Internet
11 words — < 1%

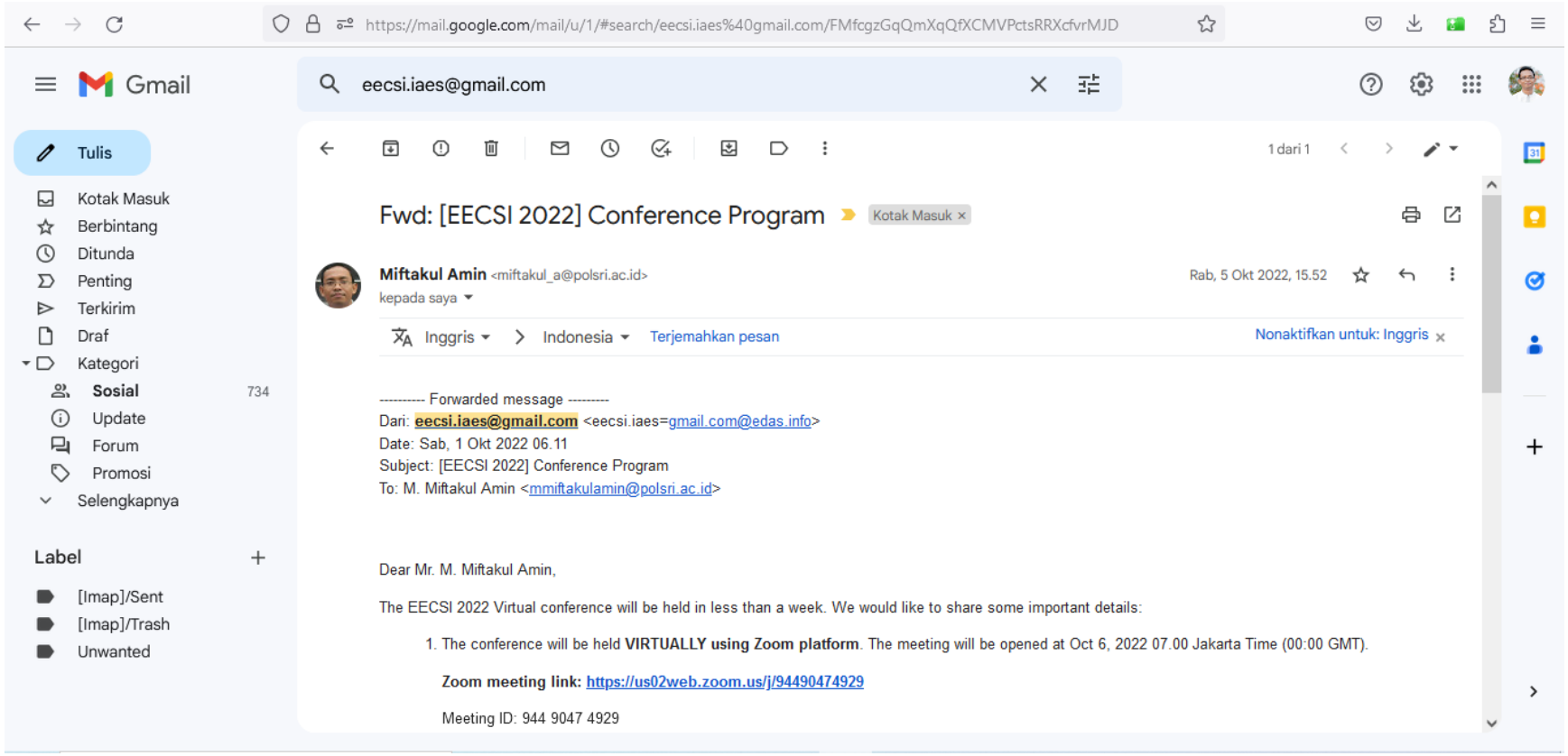
13	sweet.ua.pt Internet	10 words — < 1%
14	www.coursehero.com Internet	10 words — < 1%
15	Weifeng Su. "Record Matching over Query Results from Multiple Web Databases", IEEE Transactions on Knowledge and Data Engineering, 2009 Crossref	9 words — < 1%
16	digital.library.unt.edu Internet	9 words — < 1%
17	M. Octaviano Pratama, Pamela Kareen, Ermatita. "Building Indonesian Music Dataset: Collection and Analysis", 2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, 2021 Crossref	8 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF


https://edas.info/showSelf.php
EECSI2022 CONFERENCE
Jakarta, Indonesia
EECSI 2022 Home Register Travel grants My... Help
My... »My profile
M. Miftakul Amin Edit Conflicts List email received
Name Mr. M. Miftakul Amin
EDAS identifier 1528867
Type academia
Affiliation Politeknik Negeri Sriwijaya, Indonesia
Email addresses
Drag to reorder
Email address
Valid? (Check)
Delete Edit
mmiftakulamin@polsri.ac.id
Oct 5, 2022 07:24 America/New_York
mmiftakulamin@gmail.com
Aug 15, 2022 21:06 America/New_York


https://edas.info/editPaper.php?m=1570840744
EECSI2022
CONFERENCE
Jakarta, Indonesia
EECSI 2022 Home Register Travel grants My... Help
Edit paper #1570840744: Decision Support System using Weighting Similarity Model for Constructing Ground-Truth Dataset
Edit paper #1570840744: Decision Support System using Weighting Similarity Model for Constructing Ground-Truth Dataset
Authors can only change the paper category, change the paper topics, change the authors, change the paper title, change abstract, change the keywords, view manuscript similarity rating and report, view manuscript similarity rating, withdraw the paper after deadline, upload manuscripts (before deadline), view session information, read reviews (beyond the reviews person has written); for reviewer, after completing own reviews; question visibility and view review attachment after the deadline. For other changes, contact the conference chair.
Title of paper
Decision Support System using Weighting Similarity Model for Constructing Ground-Truth Dataset


https://edas.info/showPaper.php?m=1570840744
Personal notes
You are the creator and an author for this paper.
Reviews
2 Review extendeds
Review 1
Novelty and Contribution
Paper Presentation
Recommendation
Average (2)
Acceptable (3)
Weak Accept (4)
Detailed Comments (Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.)
Introduction:
Authors are suggested to highlight the main issue/problem presented in the manuscript. Problems are not clear enough. Therefore, authors can only proposed method to solve the problems
Method: clear enough and fair
Review 2
Novelty and Contribution
Paper Presentation
Recommendation
Average (2)
Acceptable (3)
Weak Accept (4)
Detailed Comments (Please provide detailed comments that will be helpful to the TPC for assessing the paper. Also provide feedback to the authors.)
1. The paper needs language improvement (eg. proofread),as many grammatical problems and typos are found.
2. The research intends to find duplicates on scientific publication database. However, it seems that multi-author paper may be detected as duplication. Please address this issue.
3. The reference list needs to be rechecked and revised, as many are without proper paper identification.
EDAS at alpha for 2001:448a:10e7:1f01:205d:359c:ec51:6a5d (Sat, 18 Mar 2023 10:03:28 -0400 EDT) [User 1528867 using Win10.Firefox 111.0.0.327/3.332.s] Request help


https://mail.google.com/mail/u/1/#search/eecsi.iaes%40gmail.com/FMfcgzGqQmXqQfXCMVPctsRRXcfvrMJD
Gmail
eecsi.iaes@gmail.com
Tulis
Kotak Masuk
Berbintang
Ditunda
Penting
Terkirim
Draf
Kategori
Sosial 734
Update
Forum
Promosi
Selengkapnya
Label
[Imap]/Sent
[Imap]/Trash
Unwanted
1 dari 1
Fwd: [EECSI 2022] Conference Program
Kotak Masuk ×
Miftakul Amin <miftakul_a@polsri.ac.id>
kepada saya
Rab, 5 Okt 2022, 15.52
Inggris
Indonesia
Terjemahkan pesan
Nonaktifkan untuk: Inggris
---------- Forwarded message ---------
Dari: eecsi.iaes@gmail.com <eecsi.iaes=gmail.com@edas.info>
Date: Sab, 1 Okt 2022 06.11
Subject: [EECSI 2022] Conference Program
To: M. Miftakul Amin <mmiftakulamin@polsri.ac.id>
Dear Mr. M. Miftakul Amin,
The EECSI 2022 Virtual conference will be held in less than a week. We would like to share some important details:
1. The conference will be held VIRTUALLY using Zoom platform. The meeting will be opened at Oct 6, 2022 07.00 Jakarta Time (00:00 GMT).
Zoom meeting link: https://us02web.zoom.us/j/94490474929
Meeting ID: 944 9047 4929