



INTERNATIONAL CONFERENCE

ON COMMUNICATION, COMPUTING AND NETWORKING (ICCCN-2017)

CERTIFICATE OF PARTICIPATION

This is to certify that Dr./Mr./Ms. Dolly Mittal has participated in the International Conference organized by Department of Computer Science & Engineering held at National Institute of Technical Teachers Training & Research (NITTTR), Chandigarh, India during 23-24 March, 2017. He/She has participated/~~presented~~ the paper titled Entity

Resolution using Multiple Blocking Keys and Hybrid Similarity.

Dr. Rakesh Kumar
Coordinator

Conference Chair

Dr. P. K. Tulsi
Patron

INDEX

Volume I

Paper ID	TRACK 1: Intelligent Computing Invited Speaker & Session Chair: Prof. Jitender K Chhabra, NIT Kurukshetra	Page No.
ICCCN-17-109	Data Acquisition based Web Scrapping Algorithm for Extraction of Data Sets from Patent Portal Gaurav Gupta, Neelesh Kumar, Indu Chhabra	1-6
ICCCN-17-126	Bus Arrival Detection using Bluetooth Low Energy Ankit Kumar, Abhinav Mathur, Akhilesh Kumar, Naveen Aggarwal	7-13
ICCCN-17-138	Accomplishing Granularity in Android mobile operating system without OS Tweaking Akshay Bhardwaj, A J Singh	14-21
ICCCN-17-143	Classification Techniques For Sentiment Analysis On Twitter Data Manpreet Kaur, Sheenam Malhotra	22-27
ICCCN-17-145	Comparative Study of Various Techniques for Predictive Analysis in Data Mining Shweta, Sheenam Malhotra	28-34
ICCCN-17-159	Comparative Analysis of SVM and ANN based Web Page Classification Amit Gupta, Rajesh Bhatia	35-41
ICCCN-17-161	A Comparative Analysis of Driver Drowsiness Detection System by Using Soft Computing Sapna Bhagat, V.K. Banga	42-46
ICCCN-17-173	Technological Instances of AI in Agriculture Sector Simran Chawla, Mandeep K. Chawla	47-51
ICCCN-17-174	A Proposed Approach for the Classification of Leukemia Using Artificial Bee Colony Optimization Technique and Neural Networks Rudrani Sharma, Rakesh Kumar	52-57
ICCCN-17-187	A review on Clustering Techniques using Hybrid K-Means Algorithm Navreet Kaur, Shruti Aggarwal	58-66
ICCCN-17-197	Machine Learning Solutions for an Autonomous Vehicle Jagannath Aghav, Rinku Nemade, Apoorva Nitsure	67-74
ICCCN-17-240	A Review on Acoustic Vehicular Classification Shikha Soni, Naveen Aggarwal, Dinesh Vijj, Amit Doegar	75-83
ICCCN-17-241	Application of Advance Heuristics Algorithms: Perspectives and Challenges Deep Singh Malhi, Harpreet Singh	84-91
ICCCN-17-248	A Review on Machine Learning Algorithms for Big Data Optimization Geetika Goyal, Nitesh Sharma	92-101
ICCCN-17-251	An Enhanced Approach for Movie Recommender System using Association Rule Mining Mariya Khurshid, Rakesh Kumar, Shano Solanki	102-105
ICCCN-17-258	Data Mining an Essential Tool for Intrusion Detection System: A Survey Sachin Bhardwaj, Arvind Lal	106-110

ICCCN-17-269	Hydrophobicity Detection in Dielectric Nanostructured Thin films using Artificial Neural Network Vikramaditya Dave, Ramesh Chandra	111-113
ICCCN-17-279	Gustatory Based EEG Signal Classification Using Neural Network C. KalyanaSundaram, P.Marichamy, R.R.Devu	114-121
ICCCN-17-211	Prediction of Liver Disease at early stage using data mining algorithm in Bioinformatics Rupinderjeet Kaur, Amritpal Kaur	122-128
TRACK 2: Parallel & Distributed Computing Invited Speaker & Session Chair: Prof. Sanjeev Sofat, PEC University of Technology, Chandigarh		
ICCCN-17-179	Implementation of Big Data Concerned with Elections using Map-Reduce as Novel Mining Algorithm Amandeep Kaur, Gagandeep Jagdev	129-135
ICCCN-17-180	Concentrating on Practicing Big Data in Retail Sector using Hadoop framework Ramandeep Kaur, Gagandeep Jagdev	136-143
ICCCN-17-186	Security of Big Data in Cloud: A Review Bhawna Jain, Tarunpreet Bhatia, A.K.Verma	144-150
ICCCN-17-193	A literature Survey on Response Time based Load Balancing Algorithms Aanjey Mani Tripathi, Sarvpal Singh	151-159
ICCCN-17-198	Containerization: Light Weight Virtualization Kriti Ohri	160-164
ICCCN-17-202	A Review on Job Scheduling Strategies to attain Energy Efficiency in Federated Grid System Jagroop Kaur, Amit Chhabra	165-170
ICCCN-17-215	Exploring Virtual Machine Migration Implementation for Cloud Computing Aditya Bhardwaj, Yashveer Yadav, C. Rama Krishna	171-177
ICCCN-17-217	Energy evaluation of Discrete Symbiotic Organisms Search algorithm for task scheduling in a cloud environment Megha Sharma, Amandeep Verma	178-183
ICCCN-17-220	Survey on Retrival and Updation Techniques over Encrypted Outsourced Cloud Data Gagan, C. Rama Krishna, Rohit Handa	184-193
ICCCN-17-223	A Novel Review on Ranked Keyword Searching over Outsourced Encrypted Cloud Data Vandana Saini, C Rama Krishna	194-201
ICCCN-17-226	Workflow Scheduling in Cloud Environment: A Comprehensive Review, Open Issues and Future Research Directions Gursleen Kaur, Mala Kalra	202-211
ICCCN-17-236	A Study on Virtualization in Green Cloud Anjum Mohd Aslam, Mala Kalra, Sarabjeet Singh	212-219
ICCCN-17-237	A Review of Fault-Tolerant Workflow Scheduling Techniques Urvashi Nag, Mala Kalra, Sarbjeet Singh	220-225
ICCCN-17-242	Muti-Objective Optimization Algorithms for Scheduling: A Literature Review Deep Singh Malhi, Harpreet Singh	226-232
ICCCN-17-243	Survey on Energy Efficient Techniques for Green Cloud Computing Raksha Kiran Karda, Mala Kalra, Sarabjeet Singh	233-240

ICCCN-17-245	Analysing Techniques Used For Achieving Energy Efficiency in Cloud Computing Harpreet Singh, Sandeep Sharma	241-250
ICCCN-17-246	Search with Context Sensitive Synonyms: In Multi Keyword Ranked Search over the Encrypted Cloud Data AnuKhurana, C. Rama Krishna, Navdeep Kaur	251-258
ICCCN-17-254	Survey of Various Task Scheduling Algorithms in Cloud Computing Aditi Jain, Raj Kumari	259-263
ICCCN-17-285	Security Issues in Cloud: A Review Deepti Tara, Savneet Bedi	264-268
ICCCN-17-155	Privacy-Preserving Multi-Keyword Search Supporting Synonym Query over Encrypted Data Rohit Handa, C. Rama Krishna, Naveen Aggarwal	269-277
ICCCN-17-156	Client Server Based Security Model for Private Cloud Environment Yashveer Yadav, C. Rama Krishna	278-282
ICCCN-17-164	Wind Driven Optimization based Workflow Scheduling in Cloud Computing Poonam Singh, Maitreyee Dutta, Naveen Aggarwal	283-292
ICCCN-17-250	Cost-effective and Reliable Scheduling of Workflows in Cloud using Intelligent Water Drops Algorithm Mala Kalra, Sarbjeet Singh	293-299
TRACK 3: Wireless Networks Invited Speaker & Session Chair: Prof. A K Singh, NIT Kurukshetra		
ICCCN-17-108	AMQoS: ACO based Multi-Constrained QoS Routing Protocol for Wireless Sensor Networks Tarunpreet Kaur, Dilip Kumar	300-306
ICCCN-17-123	Performance Evaluation of Chain Based Routing Technique with Mobile Sink for Wireless Sensor Networks Shailesh Kumar, Surinder Singh Khurana	307-313
ICCCN-17-147	A Collaborative Intrusion Detection System For Vehicular Ad hoc Networks Shagun Aggarwal, Sukhpreet Kaur, Gaganpreet Kaur, Rasmeet S. Bali	314-322
ICCCN-17-153	Proposed Architecture to Improve SLReduct Framework for Stress Detection using Mobile Phone Sensing Mechanism in Wireless Sensor Network Prabhjot Kaur, Sheenam Malhotra	323-329
ICCCN-17-168	Cooperative Game Theory for Residual Energy Based Load Balancing In Wireless Sensor Networks Gurwinder Singh, Surender Kumar	330-337
ICCCN-17-182	Performance Analysis of Routing Protocols in Wireless Networks Pushpender Sarao, Sanjay Pachauri, Neelam Yadav	338-341
ICCCN-17-183	A Brief Survey of Routing Protocols and an Optimal Backup Channel Routing proposal for Cognitive Radio Ad Hoc Networks Pooja Bansal, Shiraz Khurana, Prof. (Dr.) K.K. Paliwal	342-349
ICCCN-17-184	Road Map Generation for VANET Simulation Ravneet Kaur, Meenu Khurana, Rishu Chhabra	350-355
ICCCN-17-191	Review on Flexible and Secure Access to Medical Data in Wireless Sensor Networks Navneet Kaur, Sheenam Malhotra	356-359

ICCCN-17-205	An Emergency Response Approach for Wireless Industrial Sensor Networks Rajesh Kumar Verma, Joy Lal Sarkar, Rajani Trivedi, Chhabhi Rani Panigrahi, Bibudhendu Pati	360-368
ICCCN-17-206	Clustering Algorithms for VANETs : A Survey Geetanjali Goyal, Trilok C. Aseri, Sudesh Rani	369-372
ICCCN-17-210	Integrated Framework of IoT-WSN for Smart Cities Shalli Rani, Sandeep Verma	373-378
ICCCN-17-214	A Low-Cost ZigBee Based Temperature and Humidity Acquisition System for Research and Industrial Applications Vidhu Shekhar Pandey, Deepak Sharma, Awdhesh Kumar Shukla, Sachin Tyagi, Manoj Kumar Nayak, Priyanka Bajaj	379--385
ICCCN-17-228	Heterogeneity analysis of clustering based data collection techniques for WSN Sukhwinder Sharma, Rakesh Kumar Bansal, Savina Bansal	386-391
ICCCN-17-234	A Survey on Congestion Control in VANETs : Rate and Range based Approaches Sudesh Rani, Deachan Choskit	392-401
ICCCN-17-255	Taxonomy of Network layer DoS Attacks in Wireless Sensor Networks Navjot Sidhu, Monika Sachdeva	402-408
ICCCN-17-167	Health Monitoring using Cloud-Based Wireless Body Area Networks: A Survey Roopali, Rakesh Kumar	409-415

Volume II

Paper ID	TRACK 4: Computer Network and Security Invited Speaker & Session Chair: Mr. Ch A. S. Murty, C-DAC, Hyderabad	Page No.
ICCCN-17-101	A Brief Review of Methods for Prediction of Flyrock in Blasting Navdeep Kumar, Balmukund Mishra, Vikram Bali, K.K Paliwal	416-420
ICCCN-17-110	A Survey of various IP-Traceback Approaches Amrita Saini	421-425
ICCCN-17-114	Modified Otway-Rees And Needham-Schroeder Protocol For Session Key Distribution And Mutual Authentication in Distributed Computing Environment Rahila Khan, C. Rama Krishna	426-434
ICCCN-17-116	Design of Secure Data Communication for Seismic Alert System Satish Kumar, Pawan Kapur, Renu Vig	435-442
ICCCN-17-141	Performance Evaluation of Multi-Path TCP under Diversified Networks Anurag Jagetiya, C. Rama Krishna	443-448
ICCCN-17-142	Polymorphic Shellcode Detection Using Proposed SSpeaking Model Implemented through Designed Intrusion Detection Prevention System Policy Saurabh Kumar, Anup Girdhar, C. Rama Krishna	449-454
ICCCN-17-163	ARP Poisoning Detection and Prevention Techniques: A Survey Vidya Srivastava, Muzammil Hasan, Dayashankar Singh	455-461
ICCCN-17-165	A Review: Comparative Analysis of Different Methods Used to Detect Malicious Accounts in Social Web Application Anamika Saini, Sanjeev Sofat, Monika Singh	462-471
ICCCN-17-169	A Comparative Study of Variants of RPL Protocol for Low Power and Lossy Networks Sonam Goyal, Trilok C. Aseri	472-475
ICCCN-17-170	A Comparative Analysis of Dynamic Malware Analysis Tools Chahak, A.K Verma	476-482
ICCCN-17-175	Deep learning for Real Time Collision Detection and Avoidance Jagannath Aghav, Poorwa Hirve, Mayura Nene	483-488
ICCCN-17-190	A New Approach to Punjabi Text Steganography using Naveen Toli Arun Kumar, Amandeep Kaur	489-495
ICCCN-17-208	Survey of Variants of the Objective Functions in Routing Protocol for Low Power and Lossy Networks Sonia Uttreja, Trilok C. Aseri	496-499
ICCCN-17-222	An Effective Authentication Scheme for Hybrid Cloud Using Elliptic Curve Cryptography and SHA-512 Monika Gogna, C Rama Krishna	500-506
ICCCN-17-227	Throughput and Delay Performance of Cooperative TDMA MAC Protocol for Cognitive Networks Tadavarty Sriram, Thatakam Pooja, Sushma Chakravadhanula and B. N. Bhandari	507-514
ICCCN-17-230	Audio Steganography: A Review Sakshi Gupta, Deepti Dhingra, K.K.Paliwal	515-520
ICCCN-17-252	Unknown Malware Detection with Static Features Using Machine Learning : A Brief Survey Sanjay Sharma, C. Rama Krishna	521-527

TRACK 5: Digital Signal and Image Processing Invited Speaker & Session Chair: Prof. Renu Vig, UIET, Panjab University, Chandigarh		
ICCCN-17-105	A Review on GSA-FODPSO-SVM Based Feature Selection Algorithm Purnima Gaba, Surjeet Singh, K.K. Paliwal	528-533
ICCCN-17-112	Implementing Three Way Bit Wise Separation of the Speech Signal Based on Element Position in the Speech Signal Divya Sharma	534-537
ICCCN-17-113	Implementing Odd and Even Element Separation for Encryption of Speech Signal using MATLAB Divya Sharma	538-542
ICCCN-17-131	MIMO-OFDM PAPR Reduction using Fusion of Techniques Raj Lakshmi Shukla, Monika Singh	543-548
ICCCN-17-146	Advancement in Adaptive Algorithm for Non- Stationary Noise Cancellation Shelly Garg, Heena gupta, Reecha sood	549-553
ICCCN-17-150	Comparative Study of Efficient Face Recognition Methods: A Literature Survey Shashi Kant Sharma, Maitreyee Dutta, Kota Solomon Raju	554-559
ICCCN-17-157	Analysis of Gradient Based Edge Detection Operators Jaspreet Kaur, Anand Sharma	560-563
ICCCN-17-185	Implementing Software Defined Radio using USRP and GNU Radio Nayan Basumatary	564-570
ICCCN-17-188	A Historical Review of Image Forgeries Savita Walia, Krishan Kumar and Deva Prasad	571-576
ICCCN-17-192	Capacity Enhancement of MIMO System over Fading Channels: A Review Shivani, Amita Soni	577-581
ICCCN-17-199	A Hybrid CEC Approach For Estimation Of Color Light Source Harpreet Kaur, Sandeep Sharma	582-588
ICCCN-17-203	A Hybrid Approach For Stereo Matching Based On Joint Trilateral Filter & Improved Mean-Shift Deepika Kumari, Kamaljit Kaur	589-594
ICCCN-17-212	Determination of Light Direction using 3D Vector Analysis of Known Shapes in Images Deva Prasad, Krishan Kumar, SavitaWalia	595-600
ICCCN-17-213	A Survey on Nature Inspired Approaches for Brain Tumor Segmentation Khyati, Amit Doegar, Mamta Juneja	601-607
ICCCN-17-232	QoS Discrepancy Impact Aware Best Fit Channel Selection and Void Filling by Burst Segmenting and Scheduling Venkata Rao Tavanam, B.N Bhandari, Karuna Sagar Dasari	608-616
ICCCN-17-244	Performance Comparison of PAPR Reduction Techniques Asra Mubeen, S. Shalini Priya, CH. Nipun, P.Sisira, B.N. Bhandari	617-623
ICCCN-17-253	Image Processing Technique to Determine Quality Aspect of Microwave Processed Paneer Prepared by Automated Press Chitranayak, M. Manjunatha, Mahesh Kumar G., Padma Priya, P.S. Minz, Amita Vairat	624-627
ICCCN-17-261	A Survey of Gender Classification Techniques using Facial Characteristics	628-635

	Navreet Kaur, Sukhwinder Singh	
ICCCN-17-262	Multi-modal Medical Image Fusion: A Survey Jaskaranveer Kaur, Sukhwinder Singh, Savita Gupta	636-642
ICCCN-17-265	Design of N-Bit Fault Tolerant Division Unit using New Reversible Gates in QCA A. Kamaraj, P. Marichamy, J. Dhivya Bharathi	643-650
ICCCN-17-278	High Performance Interference Reduction in DS-CDMA System using Hybrid Cancellation method with Greedy Technique S.Ramkumar, J.Vijayalakshmi, V.Dinesh	651-655
ICCCN-17-282	PAPR Evaluation of Wavelet based OFDM for Underwater Acoustic Communications Naresh Kumar, B. S. Sohi	656-662
ICCCN-17-286	Review of Various Techniques Used in Image Enhancement in Vehicles Detection Alisha Sharma, Sandeep Sharma	663-670
TRACK 6: Data Science		
Invited Speaker & Session Chair: Prof. R. B. Patel, Chandigarh College of Engg. & Tech., Chandigarh		
ICCCN-17-120	SnapBook – Design & Development of an Interactive Learning Device for Kids P.Sai Sheela, Rashika Joshi, Jasleen Kaur, SRN Reddy	671-677
ICCCN-17-125	Pi Attendance: Design and Development of Biometric Portable Attendance System Deepika Wadhwa, Jasleen Kaur, SRN Reddy	678-682
ICCCN-17-132	Intel Genuino 101: Case Study of Intel Curie Computing Platform for IoT Applications Nisha Sharma, Zeenat Shareef, S.R.N Reddy	683-689
ICCCN-17-151	A Framework for Data Quality Management in ERP implementation Ranjeet Kaur, Rakesh Kumar	690-695
ICCCN-17-152	Reliability Analysis of an Augmented Baseline Network Using Markov Model Shally Bansal, Shalli Rani	696-702
ICCCN-17-172	Software Component Identification from Object-Oriented Code Using Structural & Semantic Dependencies Amit Rathee, Jitender Kumar Chhabra	703-707
ICCCN-17-177	Entity Resolution using Multiple Blocking keys and Hybrid Similarity Dolly Mittal, Emmanuel S Pill, M.C. Govil	708-714
ICCCN-17-178	A Review on Sentiment Analysis Shruti Gupta, Ashutosh Pandey, Vikram Bali	715-720
ICCCN-17-239	Microsoft Windows Azure Technology: An Overview Rupali J. Dhabarde	721-724
ICCCN-17-247	The Need for Integrating Cloud Computing and Internet of Things: A Review Ramneek Kaur, C. Rama Krishna	725-731
ICCCN-17-270	Design of a Body Matched Bowtie Antenna for Biomedical Applications R. Rajkumar, P. Marichamy	732-735
ICCCN-17-271	A Review on Medical Image Fusion Aspects and Techniques Gurpreet Kaur, Renu Vig, Sukhwinder Singh	736-741
ICCCN-17-277	Spectrum Access in Cognitive Radio Networks: A Comparative	742-748

	Study Bhoopendra Kumar, Sanjay Kumar Dhurandher	
ICCCN-17-119	Big Data Analytics for Quality Improvement in Higher Education P S Grover	749-752
ICCCN-17-139	Demonstration of IoT on IBM Bluemix: Temperature Sensor Application as a Case Study Anjum Mohd Aslam, Raksha Kiran Karda, Aditya Bhardwaj, C. Rama Krishna	753-760
ICCCN-17-209	Cloud Based E Governance: Benefits and Challenges Deepti Sehrawat, Deepti Sindhu, Palak	761-768
ICCCN-17-231	Insights from Big Data Punita, Amit Deogar	769-775

Entity Resolution using Multiple Blocking keys and Hybrid Similarity

Dolly Mittal^{*†}, Emmanuel S Pilli^{*}, M.C. Govil^{*}

^{*}Department of Computer Science & Engg., Malaviya National Institute of Technology Jaipur

[†]Department of Computer Science Engg., Swami Keshvanand Institute of Technology Jaipur,
2014pcp5269@mnit.ac.in, espilli.cse@mnit.ac.in

Abstract—Entity Resolution plays a significant role in data integration task and defined as identifying entities referring to the same real world object. The standard ER process compare each entity with all other entities, resulting into a complexity of $O(n^2)$, which is inefficient for large datasets. The approach used to improve efficiency of this process is to reduce the search space by adopting blocking techniques based on blocking key. But the concept of single blocking key provides incomplete results which can be improved using the concept of multiple blocking keys, as the probability of identifying two similar entities increased if they share multiple blocking keys. This approach decreases the chance of missing a similar entity pair but suffers from duplicated pair matching tasks.

Further, parallel and distributed execution framework MapReduce is used for efficient execution of this data intensive task of entity resolution. Efficient partitioning technique is used to remove the limitations of skewed dataset and matching task are evenly distributed among all the reducer.

So in this paper we propose an efficient MapReduce algorithm for entity resolution in bibliographic dataset using multiple blocking keys. Also, hybrid similarity measure is used to determine the similarity between two entities. We used a bibliographic dataset in our experiment to show that our approach is efficient as well as scalable.

Keywords—Map-Reduce, Entity Resolution, Blocking, Hybrid similarity

I. INTRODUCTION

Entity Resolution is the task of identifying entities referring to the same real world

object. For example to find duplicate entries of customer in an enterprise database or to match product offers for price comparison portals. Massive amount of data is being generated by many real world domains like social networking, pharmaceutical and healthcare, telecom- munications, E-Commerce websites. For Efficient processing, management and analysis of large data collections various novel techniques are required [1]. That is why fields like data warehousing and data mining have gained popularity in both academia and industry.

Entity Resolution also known as entity/object matching, data deduplication, record linkage is a complex problem and has a significant impact on data quality and data integration. ER is an important step of data integration process. While entity match- ing is concerned with identifying and matching individual entities that refer to the same real world objects from several data sources, data integration is the overall process of integrating heterogeneous datasets or data repositories to provide a unified and clear view of the available data. Entities used in ER mainly refer to people, such as patients, customers, taxpayers, or travellers, but they are also applicable for consumer goods and products, publications or citations, or businesses. Take a bibliographic database for example, In the era of digitization research publications from academia and industry are available electronically and through online databases such as IEEE Xplore, Google Scholar etc, Knowledge. This digitization of research sets a new direction for research scholars as they are