Advances in Intelligent Systems and Computing 1158

Sanjiv K. Bhatia · Shailesh Tiwari ·
Su Ruidan · Munesh Chandra Trivedi ·
K. K. Mishra *Editors*

Advances in Computer, Communication and Computational Sciences

Proceedings of IC4S 2019

Springer

Advances in Intelligent Systems and Computing

Volume 1158

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****

More information about this series at http://www.springer.com/series/11156

Sanjiv K. Bhatia · Shailesh Tiwari ·
Su Ruidan · Munesh Chandra Trivedi ·
K. K. Mishra
Editors

Advances in Computer, Communication and Computational Sciences

Proceedings of IC4S 2019

Springer

Editors
Sanjiv K. Bhatia
Department of Mathematics
and Computer Science
University of Missouri–St. Louis
Chesterfield, MO, USA

Su Ruidan
Shanghai Advanced Research Institute
Pudong, China

K. K. Mishra
Computer Science Engineering Department
Motilal Nehru National Institute
of Technology
Allahabad, Uttar Pradesh, India

Shailesh Tiwari
Computer Science Engineering Department
ABES Engineering College
Ghaziabad, Uttar Pradesh, India

Munesh Chandra Trivedi
National Institute of Technology Agartala
Agartala, Tripura, India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-15-4408-8 ISBN 978-981-15-4409-5 (eBook)
https://doi.org/10.1007/978-981-15-4409-5

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The IC4S is a major multidisciplinary conference organized with the objective of bringing together researchers, developers and practitioners from academia and industry working in all areas of computer and computational sciences. It is organized specifically to help computer industry to derive the advances of next-generation computer and communication technology. Researchers invited to speak will present the latest developments and technical solutions.

Technological developments all over the world are dependent upon globalization of various research activities. Exchange of information and innovative ideas is necessary to accelerate the development of technology. Keeping this ideology in preference, the International Conference on Computer, Communication and Computational Sciences (IC4S 2019) has been organized at Mandarin Hotel Bangkok, Bangkok, Thailand, during 11–12 October 2019.

This is the third time the International Conference on Computer, Communication and Computational Sciences has been organized with a foreseen objective of enhancing the research activities at a large scale. Technical Program Committee and Advisory Board of IC4S include eminent academicians, researchers and practitioners from abroad as well as from all over the nation.

In this book, selected manuscripts have been subdivided into various tracks named—Advanced Communications and Security, Intelligent Hardware and Software Design, Intelligent Computing Techniques, Web and Informatics and Intelligent Image Processing. A sincere effort has been made to make it an immense source of knowledge for all and includes 91 manuscripts. The selected manuscripts have gone through a rigorous review process and are revised by authors after incorporating the suggestions of the reviewers.

IC4S 2018 received around 490 submissions from around 770 authors of 22 different countries such as USA, Iceland, China, Saudi Arabia, South Africa, Taiwan, Malaysia, Indonesia, Europe and many more. Each submission has been gone through the plagiarism check. On the basis of plagiarism report, each submission was rigorously reviewed by atleast two reviewers with an average of 2.4 per reviewer. Even some submissions have more than two reviews. On the basis of these reviews, 91 high-quality papers were selected for publication in this proceedings volume, with an acceptance rate of 18.57%.

We are thankful to the keynote speakers—Prof. Shyi-Ming Chen, IEEE Fellow, IET Fellow, IFSA Fellow, Chair Professor in National Taiwan University of Science and Technology, Taiwan, and Prof. Maode Ma, IET Fellow, Nanyang Technological University, Singapore, to enlighten the participants with their knowledge and insights. We are also thankful to delegates and the authors for their participation and their interest in IC4S 2019 as a platform to share their ideas and innovation. We are also thankful to the Prof. Dr. Janusz Kacprzyk, Series Editor, AISC, Springer, for providing guidance and support. Also, we extend our heartfelt gratitude to the reviewers and Technical Program Committee Members for showing their concern and efforts in the review process. We are indeed thankful to everyone directly or indirectly associated with the conference organizing team leading it towards the success.

Although utmost care has been taken in compilation and editing, however, a few errors may still occur. We request the participants to bear with such errors and lapses (if any). We wish you all the best.

Editors

Bangkok, Thailand

Sanjiv K. Bhatia
Shailesh Tiwari
Munesh Chandra Trivedi
K. K. Mishra

About This Book

With advent of technology, intelligent and soft computing techniques came into existence with a wide scope of implementation in engineering sciences. Nowadays, technology is changing with a speedy pace and innovative proposals that solve the engineering problems intelligently are gaining popularity and advantages over the conventional solutions to these problems. It is very important for research community to track the latest advancements in the field of computer sciences. Keeping this ideology in preference, this book includes the insights that reflect the *Advances in Computer and Computational Sciences* from upcoming researchers and leading academicians across the globe. It contains the high-quality peer-reviewed papers of 'International Conference on Computer, Communication and Computational Sciences (IC4S-2019)', held during 11–12 October 2019 at Mandarin Hotel Bangkok, Bangkok, Thailand. These papers are arranged in the form of chapters. The content of this book is divided into five broader tracks that cover variety of topics. These tracks are: *Advanced Communications and Security, Intelligent Hardware and Software Design, Intelligent Computing Techniques, Web and Informatics and Intelligent Image Processing*. This book helps the perspective readers' from computer and communication industry and academia to derive the immediate surroundings developments in the field of communication and computer sciences and shape them into real-life applications.

Contents

Intelligent Image Processing

About the Editors

Sanjiv K. Bhatia works as a Professor of Computer Science at the University of Missouri, St. Louis, USA. His primary areas of research include image databases, digital image processing, and computer vision. In addition to publishing many articles in these areas, he has consulted extensively with industry for commercial and military applications of computer vision. He is an expert on system programming and has worked on real-time and embedded applications. He has taught a broad range of courses in computer science and has been the recipient of the Chancellor's Award for Excellence in Teaching. He is also the Graduate Director for Computer Science in his department. He is a senior member of ACM.

Shailesh Tiwari currently works as a Professor at the Department of Computer Science and Engineering, ABES Engineering College, Ghaziabad, India. He is an alumnus of Motilal Nehru National Institute of Technology Allahabad, India. His primary areas of research are software testing, implementation of optimization algorithms, and machine learning techniques in software engineering. He has authored more than 50 publications in international journals and the proceedings of leading international conferences. He also serves as an editor for various Scopus, SCI, and E-SCI-indexed journals and has organized several international conferences under the banner of the IEEE and Springer. He is a senior member of the IEEE and a member of the IEEE Computer Society.

Su Ruidan is currently an Assistant Professor at Shanghai Advanced Research Institute, Chinese Academy of Sciences. He has completed his Ph.D. from Northeastern University in 2014. His research areas include machine learning, computational intelligence, software engineering, data analytics, system optimization, multi-population genetic algorithm. Dr. Su has served as Editor-in-Chief of the journal "Journal of Computational Intelligence and Electronic Systems" during 2012–2016. He has published more than 20 papers in international journals.

Dr. Munesh Chandra Trivedi is currently working as Associate Professor, Department of Computer Science & Engineering, National Institute of Technology, Agartala (Tripura). He worked as Dean Academics, HoD & Associate Professor (IT), Rajkiya Engineering College with additional responsibility of Associate Dean UG Programs, Dr. APJ Abdul Kalam Technical University, Lucknow (State Technical University). He was also the Director (In charge) at Rajkiya Engineering College, Azamgarh. He has a very rich experience of teaching the undergraduate and postgraduate classes in Government Institutions as well as prestigious Private institutions. He has 11 patents in his credit. He has published 12 text books and 107 research papers publications in different International Journals and in Proceedings of International Conferences of repute. He has also edited 21 books of the Springer Nature and also written 23 book chapters for Springer Nature. He has received numerous awards including Young Scientist Visiting Fellowship, Best Senior Faculty award, outstanding Scientist, Dronacharya Award, Author of Year and Vigyan Ratan Award from different national as well international forum. He has organized more than 32 international conferences technically sponsored by IEEE, ACM and Springer's. He has also worked as Member of organizing committee in several IEEE international conferences in India and abroad. He was Executive Committee Member of IEEE UP Section, IEEE computer Society Chapter India Council and also IEEE Asia Pacific Region-10. He is an active member of IEEE Computer Society, International Association of Computer Science and Information Technology, Computer Society of India, International Association of Engineers, and life member of ISTE.

K. K. Mishra is currently working as an Assistant Professor at the Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, India. He has also been a Visiting Faculty at the Department of Mathematics and Computer Science, University of Missouri, St. Louis, USA. His primary areas of research include evolutionary algorithms, optimization techniques and design, and analysis of algorithms. He has also authored more than 50 publications in international journals and the proceedings of leading international conferences. He currently serves as a program committee member of several conferences and an editor for various Scopus and SCI-indexed journals.

Visualizing Academic Experts on a Subject Domain Map of Cartographic-Alike

Diana Purwitasari, Rezky Alamsyah, Dini Adni Navastara, Chastine Fatichah, Surya Sumpeno, and Mauridhi Hery Purnomo

Abstract Visualizing bibliographic information aids academician users to gain insights into science mapping, and then to define the next research plans. This paper focused on expert visualization to make users utilizing their cognitive skills to comprehend science mapping by exploring experts and domain expertise. To address the comprehension problem, we represented the knowledge domain and the involving players or academic experts in a visual approach of cartographic-alike. First, to generate a base map of standardized knowledge domains, we identified semantic relatedness through word embedding on collected metadata texts of articles according to Scopus subject areas. Then, the expert coordinates were obtained after transforming article metadata with the base map and the articles were labeled with subject domains. To make it cartographic-alike, subject domain color on the map was set, where darker areas indicated more experts had interests in the particular subjects, while blended colors demonstrated mixed subjects. The experiments required two semi-manually collected datasets of Domain Data and Researcher Data in the forms of Scopus metadata. Our findings on the embedding process showed that

D. Purwitasari (✉) · S. Sumpeno · M. H. Purnomo
Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
e-mail: diana@if.its.ac.id

S. Sumpeno
e-mail: surya@ee.its.ac.id

M. H. Purnomo
e-mail: hery@ee.its.ac.id

D. Purwitasari · R. Alamsyah · D. A. Navastara · C. Fatichah
Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
e-mail: rezkyal2@gmail.com

D. A. Navastara
e-mail: dini_navastara@if.its.ac.id

C. Fatichah
e-mail: chastine@cs.its.ac.id

S. Sumpeno · M. H. Purnomo
Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

S. K. Bhatia et al. (eds.), *Advances in Computer, Communication and Computational Sciences*, Advances in Intelligent Systems and Computing 1158, https://doi.org/10.1007/978-981-15-4409-5_63

labeling articles, and hence, experts gave a better performance with training on 1st tier Scopus subject areas of four domains compared to 2nd tiers of 26 sub-domains to avoid over-mixed subjects in the article contents. The visual result of the cartographic colored map had encouraged the respondents to explore the research interest of the experts. After observing the color blended map, users could be expected to initiate crossed domain collaboration plans.

Keywords Visualizing experts · Subject domain map · Word embedding · Article labeling

1 Introduction

Bibliographic information for scientific literature datasets contains scientific article metadata of title, authors, abstract, keywords, and references. With the increase of open datasets, science mapping as the process of mining and analysis of the metadata helps the user in finding emerging topics and making research plans. Since the beginning of bibliometric research, there have been many studies that introduce the visual approaches to gain insights into science mapping. Visualizing topics as a vital process in science mapping can be roughly categorized into topic content, topic relationship, and topic evolution [1]. There was a cartographic approach to display topic contents of scientific literatures from keywords and their semantic relations on the map [2]. Another visualization aspect of the topic relationship was about representing co-citations in tree-based structures [3]. However, mechanisms of zooming, searching, or filtering became challenging issues, not to mention explanation efforts to make user understands the meaning of tree structure. Map-related technique was also used with spatial autocorrelation analysis to generate a Delaunay triangulation network of keyword polygons. The map illustrated topic evolution issue through categories inspired by emerging topic statuses with the influence of semantic relatedness among keywords [4]. The visual approach also considered the article authors in the network forms [5]. Many visualization approaches described only the relations in the knowledge domain or just the relations between the involving authors or experts of the knowledge domain. Yet, combining both relationships between knowledge domain and the involving players, or defined as academic experts, in visualization is less studied.

Previous descriptions took account of visualizing topics as part of the knowledge domain in the science mapping, which often identified from article texts. Without any predefined information, clustering became common approach like an expert map generated by self-organizing map (SOM) algorithm on dataset of researches and their listed interest [6]. Notes that with other datasets, the SOM approach would have different maps. Trends of research topics are intuitively observable through the same set of subject areas. However, positions of subject domains on the map can be entirely different because article metadata as the domain source is periodically updated. For that reason, this paper proposed a system for visualizing experts on a standardized map of subject domains. The subsequent sections explained about four main processes in the proposed system and then followed with empirical experiments using metadata of local academicians manually collected from Scopus.

2 Related Works

Visualizing scientific literatures generally emphasizes on texts of title-abstract, citations of networks, authors, and metadata which includes all three parts through the mechanisms of looking up, seeking relations or finding temporal pattern [7]. Still related to over time analysis, citations combined with temporal patterns were allowing users to track long-term developments for seeing the research trends [8], whereas the networks of citations or co-authorships were utilized to predict the possibilities for activities of the researchers [9]. With topics as the focus, the concerned process included defining research areas, knowing relationships between those subject areas and then predicting subject evolution. To incorporate the authors, mapping them to the defined research areas and then identifying relationships between them before predicting future collaborations are still interesting for explorations with some supported visuals. Those information were defined as author profiles [10], although that visual did not display authors onto a map of subject areas. The map is expected self-explanatory interface to support users in understanding any information related to scientific literatures. Hence, the next proposed process here established a standardized and self-explanatory map for exhibiting experts and their research interests.

3 Proposed Methodology to Visualize Experts on Subject Domain Map

There are four main processes in Fig. 1, namely collecting article metadata of researchers and domain-related, then preparing the base map from knowledge domain according to Scopus subject area. After transforming articles in reference to the base map, the scaling process makes the visualization result displaying subject areas of the experts. Here, users were expected to utilize the proposed system, called in Indonesian language as SIPETRUK that representing features of mapping, recommending, and visualizing to explore experts and their subject domains through a map. Later, we described the context of use for SIPETRUK and other existing nationwide systems that related in supporting policies to increase the national research productivity level.

The process of preparing subject domains in Fig. 2 used dataset of article metadata from predefined subject domains of Scopus subject areas. Terms in texts of titles and abstracts from articles in Domain Data were tagged based on parts of speech because further process only required noun keywords. Preparing subject domain

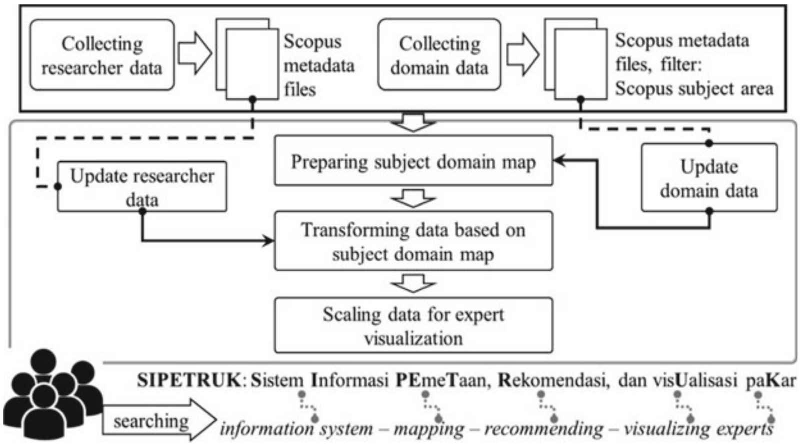


Fig. 1 System architecture for visualizing academic experts on a subject domain map of cartographic-alike

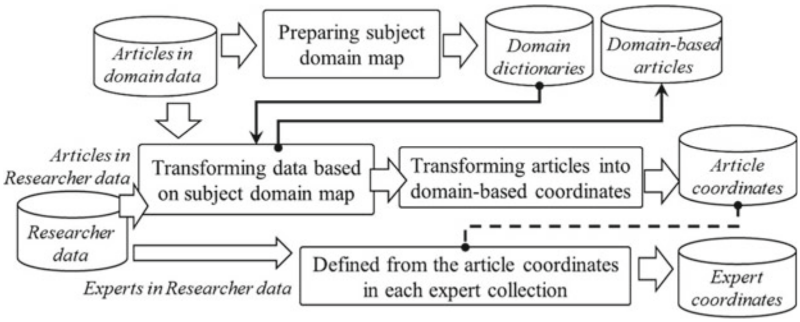


Fig. 2 Transforming metadata based on information of subject domains

applied Word2Vec word embedding using a shallow neural network [11] on those keywords. It resulted into vectors of weight values of important words in Domain Dictionaries. Vectors with closer values represented similar semantic relationship of noun keywords. Then, raw texts of title-abstract of published articles for each researcher (from Researcher Data) and Scopus recent articles based on subject areas (from Domain Data) were transformed with Domain Dictionaries. The transformation results of word embedding approach were kept as Domain-based articles, in which all represented articles became vectors with the same dimensions.

To visualize the articles on a subject domain map, all vectors had 2D transformation of t-Distributed Stochastic Neighbor Embedding (t-SNE) [12] and being generated into *x*-*y* coordinates which more suitable for user observation. Article vectors tend to be high-dimensional data because number of keywords becomes number of dimensions. Method t-SNE was used to project the article vectors to a

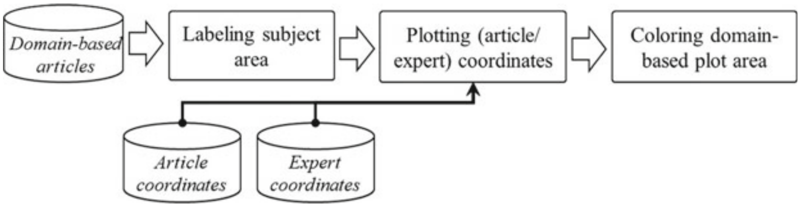


Fig. 3 Scaling in expert visualization with cartographic-alike using transformed data

low-dimensional space. It still maintained the local structure or distances between the points to remain almost the same without turning them into crowded points. Two points of articles were located closer if they had similar context, and it could be expected that both articles utilize the same keywords. The focus on the proposed system was visualizing the experts instead of articles of the experts. Consequently, after t-SNE transformation, expert coordinates were obtained through calculating the center value of article coordinates from each expert collection respectively.

Steps in Fig. 3 are displaying the expert coordinates and other meaningful information on the map, so that users could apply their cognitive skills to explore the visualization to comprehend science mapping. In the visualization, labels of Scopus subject areas were covered in the base map and became the information of experts as well. As a consequence, expert labels were established from article labels. This was similar to the process for obtaining expert coordinates from article coordinates. Given Word2Vec transformed articles in Domain-based articles, a test data of an expert article from Researcher Data was labeled by a lazy learning approach inspired from k-nearest neighbors (kNN) [13]. Noted that, a test data of an expert article was also Word2Vec transformed. If an expert had a number of articles labeled in certain subject more than a threshold value, the expert was stated to have the particular subject as his or her research interest.

For coloring the map in Fig. 3 to help user cognitive skill, the base map was set to have scaled grids. Each grid color depends on the numbers of subjects and experts. Various subjects influenced the blended aspect in coloring, while the number of experts in one subject affected the transparency aspect.

4 Datasets

For our experiments, we had two collections of articles called as Researcher Data and Domain Data from Scopus metadata (Fig. 1). Subject details of the collected articles were listed in Table 1. As the first collection of Researcher Data, we used Scopus metadata of scientific literature filtered with the keyword "computer science" for selecting researchers or lecturers in our university, which is a public institution emphasized on the study fields of scientific and engineering. The manually collected

Table 1 List of Scopus subject areas along with the labeled numbers of articles and experts

Scopus subject area (Tier I)	Scopus subject area (Tier II)	Scopus code	Number of articles	Number of faculty experts
Health Sciences-HS (± 10,000 articles) color: Red (89)	There are no auto-labeled articles of Researcher Data for Medicine (MEDI), Nursing (NURS), Veterinary (VETE) and Dentistry (DENT)			
	Health Profession	HEAL	89	FTE(6), FTIK(2)
Life Sciences-LS (± 10,000 articles) color: Green (35)	There are no auto-labeled articles of Researcher Data for Biochemistry, Genetics, and Molecular Biology (BIOC), Immunology and Microbiology (IMMU), and Neuroscience (NEUR)			
	Agricultural and Biological Sciences	AGRI	24	FTI(1)
	Pharmacology, Toxicology and Pharmaceutics	PHAR	11	FS(2)
Physical Sciences-PS (± 20,000 articles) color: Blue (2399)	There is no auto-labeled articles of Researcher Data for Chemistry (CHEM)			
	Chemical Engineering	CENG	84	FTI(2), FS(3)
	Computer Science	COMP	1373	FTE(39), FTIK(20), FMKD(10), FTI(10), FTK(2)
	Earth and Planetary Sciences	EART	70	FTI(10), FS(2), FTSLK(2), FTE(2), FMKD(1),
	Energy	ENER	322	FTE(13), FTI(9), FMKD(1), FS(1), FTIK(1), FTK(1)
	Engineering	ENGI	6	FTI(1)
	Environmental Science	ENVI	39	FS(2), FTI(2), FTI(1), FTSLK(1)
	Materials Science	MATE	62	FTI(4), FS(1), FTSLK(1)
	Mathematics	MATH	363	FTE(13), FMKD(9), FTSLK(3), FS(2), FTI(2), FTIK(2)
	Physics and Astronomy	PHYS	80	FTI(3), FMKD(2), FS(2), FTE(1)

(continued)

Table 1 (continued)

Scopus subject area (Tier I)	Scopus subject area (Tier II)	Scopus code	Number of articles	Number of faculty experts
Social Sciences-SS (± 12,000 articles) color: Black (793)	Arts and Humanities	ARTS	41	FTE(3), FTI(1)
	Business, Management and Accounting	BUSI	132	FTI(5), FTIK(3), FTSLK(3), FBMT(1)
	Decision Sciences	DECI	581	FMKD(15), FTIK(16), FTI(12), FTE(4), FTSLK(2), FTK(2)
	Economics, Econometrics and Finance	ECON	31	FMKD(3), FTIK(1)
	Psychology	PSYC	8	FTE(1)

There are no auto-labeled articles for Social Sciences (SOCI)

metadata contained a collection of text files for each researcher. One text file represented a BibTeX database file formed by a list of entries for articles published by a researcher.

The focus of this paper was to visualize scientific mapping of experts. Therefore, we did not include metadata files of researchers who have less than ten published articles. It made the dataset contained 3182 articles of 200 researcher files from eight faculties with the following abbreviations in Indonesian language.

- FTI (industrial technology) 46 experts
- FTE (electrical technology) 44 experts
- FTIK (information and communication technology) 39 experts
- FMKD (mathematics, computation, and data science) 24 experts
- FTSLK (civil, environmental, and geoengineering) 14 experts
- FS (basic science) 9 experts,
- FTK (marine technology) 8 experts, and
- FBMT (business and technology management) 1 expert

Whereas the second collection of Domain Data focused on 26 Scopus subject areas, with 51,939 bib-items, as shown in Table 1 and each subject area had ±2000 article metadata of title-abstract published from 2017 to 2018. There are two levels of categories in Scopus subject area, i.e., 1st tier subject of physical sciences has 2nd tier subjects of computer science among others. After labeling articles and then labeling experts, Table 1 shows that a researcher might have several Scopus subject areas. Labeling process was completed if the number of labeled articles in one subject ≥5. Our initial analysis showed that prolifically experts were existed in the faculties of FTI, FTE, FTIK, and FMKD. There were three prominent experts in Table 1 with h-Index at least 11 who had several interests until 5–8 domains of 2nd tier subjects.

5 Results and Discussions

Some Python packages used in the experiments were BibtexParser for parsing raw Scopus metadata, Natural Language Toolkit (NLTK) for text processing, Gensim for word embedding, Scikit-learn for labeling, Seaborn and Matplotlib for visualization, in addition to Mpld3 for bringing the visual into web browser.

5.1 Parameter Setting for Word Embedding

Word embedding with Word2Vec required to set some parameters like minimal term occurrences in the collection or distance window to check semantic relations with nearest neighbor terms. Some setting conditions with different values in Table 2 supported the process of preparing subject domain map which was necessary for transforming data based on the domain map. With Gensim package there were two approaches of common bag of words (CBOW) and Skip Gram, as well as options to use all keywords with minimal occurrences in documents or filtered by minimal Document Frequency, DF. This resulted into six conditions for word embedding of 200 dimensions.

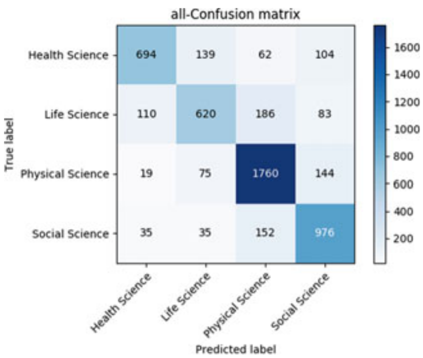
Then to observe its performance, word embedding result was used for classifying articles in Domain Data with kNN ($k = 100$) and tenfold validations. Articles in Domain Data had labels of Scopus subject areas while Research Data had not. Average precision over all 26 classes of 2nd tier subjects became the performance indicator since labeling process prioritizes true positive data. However, the precision values were quite low with less than 50%, although Skip Gram (parameter-4 Table 2) gave slightly better result compared to CBOW (parameter-6 Table 2), since Skip Gram weighting took account not randomly word usages in the sentences. Lower precision values indicated that content of the articles was mixed subjects. More keywords included within embedding process were preferable in scientific texts of mixed subjects based on the results without using DF threshold and less value in the minimal count (parameter-1 vs. parameter-4).

Table 2 Parameter values for Word2Vec word embedding used in kNN of tier-2 subjects

<table>
<tr><th>No</th><th>Word2Vec</th><th>Document Freq.</th><th>Min. Count</th><th>Window</th><th>Avg. Precision</th><th># Terms</th></tr>
<tr><td>1</td><td rowspan="2">CBOW</td><td>Use min DF = 0.1</td><td rowspan="2">10</td><td rowspan="4">5</td><td>15.15%</td><td>–</td></tr>
<tr><td>2</td><td rowspan="5">Not Use DF, based on minimal word occurrences in min. count</td><td>40.96%</td><td>19.607</td></tr>
<tr><td>3</td><td rowspan="3">Skip Gram</td><td>5</td><td>41.16%</td><td>29.674</td></tr>
<tr><td>4</td><td rowspan="3">1</td><td>43.38%</td><td rowspan="3">74.984</td></tr>
<tr><td>5</td><td>3</td><td>41.15%</td></tr>
<tr><td>6</td><td>CBOW</td><td>5</td><td>43.23%</td></tr>
</table>

Table 3 kNN classification with selected word embedding parameters of tier-1 subjects

Classes	Precision (%)	Accuracy (%)
Social Science (only 2nd tier)	75.00	81.46
	40–60	*20.5–77.5*
Physical Science (only 2nd tier)	81.00	88.08
	25–73	***3.5***–*71.85*
Life Science (only 2nd tier)	71.00	63.06
	56–75	*47–74.5*
Health Science (only 2nd tier)	81.00	69.46
	59–***91***	*65*–***80.5***
Average	78.00	77.97



Mixed subjects: Physical Science (MATE, ENGI has lowest true positive)

For example in the mixed subjects, there was an article of "*Smart meter based on time series modify and neural network for online energy monitoring*" labeled with COMP but it could be tagged as ENER as well. Noted that, both subjects belong to the same 1st tier. Therefore, with the setting condition of parameter-4 (Skip Gram, not using DF, minimal word count =1, and term distance window = 5), we classified the articles into four classes of HS, LS, PS, and SS as shown in Table 3. The results of precision (only considering true positive) and accuracy (considering both true positive and true negative) had better performance with average values of ±78%. There was an increasing of >35% compared to labeling with 2nd tier subjects as classes, which supported the mixed issue in the contents of articles as shown in the confusion matrix. The numbers of mismatched labels (not in diagonal) were quite high, and the lowest accuracy was 3.5% in ENGI subject with <10 true positive labels.

5.2 *Visualization Results*

We used two scenarios in visualizing the experts on the base map, without and with considering subject domains for coloring as shown in Figs. 4 and 5. The first visual applied color grading with density based of expert coordinates from Python package Seaborn. Positions of experts in Fig. 4 were located in Physical Sciences (a Blue gradation) especially on COMP and overlapped with Social Sciences (a Black gradation) on DECI (Table 1). There were some researches on biomedical engineering, such that several experts had another coincidence spot on Health Sciences (a Red gradation).

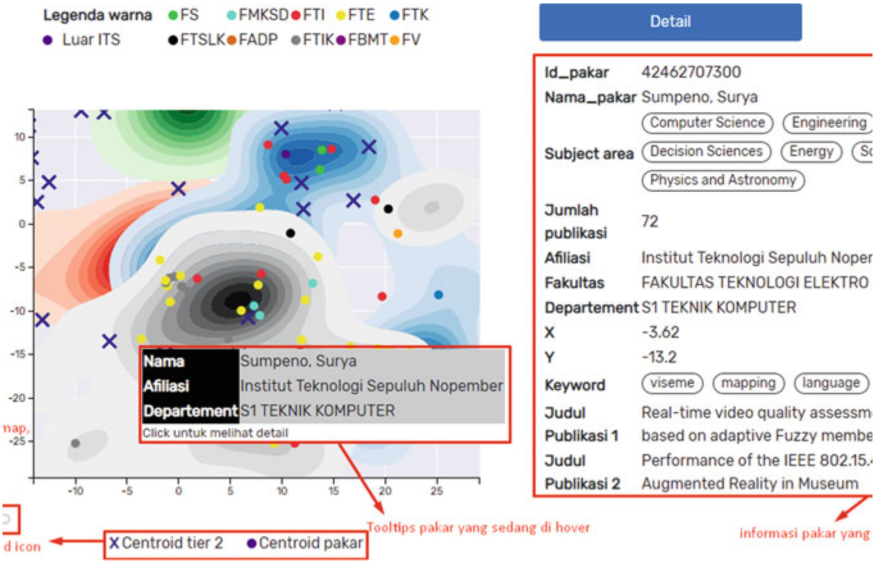


Fig. 4 Visualization without considering subject domains of the experts

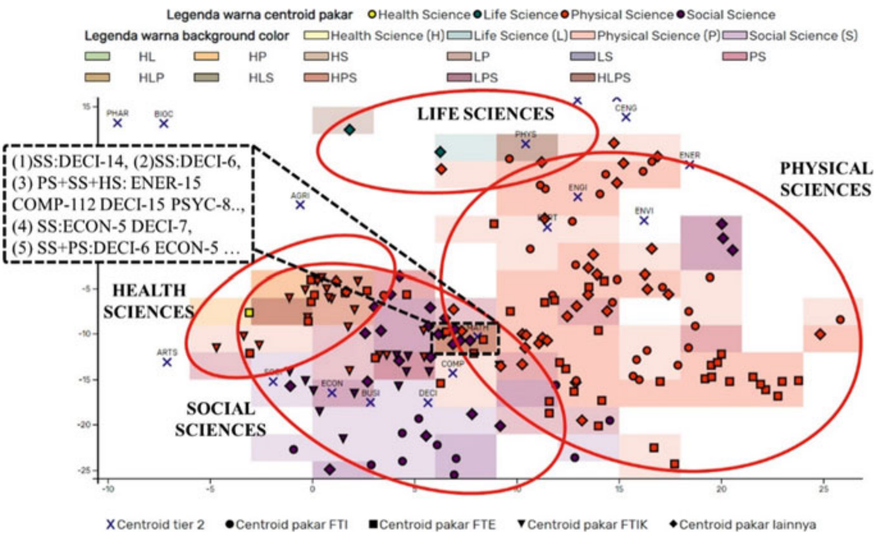


Fig. 5 Visualization with considering subject domains of the experts

The coordinates found in Life Sciences (a Green gradation) were placed in the outside area which was not surprising since Researcher Data was collected with a keyword of "computer science."

We asked 12 undergraduate students of Informatics department who were doing the graduation project in their seven or eight semesters. The questionnaire was about position suitability of ten experts or lecturers of Informatics department based on their published articles with grading values of 1–3–5. Their responds resulted into a suitability score of 3.67 and the respondents agreed that the visual aided them to find an expert.

The second scenario in Fig. 5 showed a visual with a consideration of mixed subject domains for coloring groups of experts that resulted into blended colors. Here, the blended condition examined the 1st tier subjects instead of 2nd tier subjects based on the experiment result in Table 3. The base map was divided into square grids of 3×3 units. All articles belong to the experts in one grid were considered as a collection. The number of blended colors in one grid depended on the number of the 1st tier subjects from the grid collection of articles. To ensure the blending process for cartographic-alike, the color had opacity value to make transparent effect with threshold = 20 articles. For example, if a grid had >20 articles of Physical Science (PS) then the opacity of PS was set to 25% and the grid had thicker color of PS. However, if the grid had $\leqq$ 20 articles of PS, then the opacity of PS was 12.5% and the grid color became more transparent.

Figure 5 demonstrates grids with blended colors and some gradation levels of transparency. After exploring the expert coordinates by mouse hover in the browser, coincide areas of four subject domains were identified. With one boxed area in the example, there was one square grid with at least five experts having accumulated blended colors of HS-25% + PS-25% + SS-25%; i.e., expert(1) was labeled as SS:DECI-14.

The experts were researchers from the faculties of FTE and FTIK where their research subjects often related to the keyword "computer science." Moreover, the grid example revealed that noticeable subjects of 2nd tiers among others were DECI and ECON. Both subjects had strong adjacent context to the keyword "computer science" as well. The experiments had demonstrated a subject domain map with experts on the top and having colored area in cartographic-alike because of various combinations of blend and transparency. The visualization in Fig. 5 bridged mixed subject domains with cartographic-alike approach so the users who explore it could be encouraged to strategize further collaborations.

6 Conclusions

We proposed a visual approach to display experts on a standardized subject domain map. The results of the proposed method demonstrated a visualization with a case study in the experiment. Experts of the experiment results were identified to have several related research interests as shown in blended and transparency colors on the map. Selection of subject domains was influential in defining the base map and labeling steps to produce better science mapping.

In the recent years, our government launched some systems inspired by traditional characters in our cultures, such as **Scien**ce and **T**echnology Index **A**pplication (SINTA) and **Akr**editasi **Ju**rnal **Na**sional, in Indonesian language, for maintaining a list of nationally accredited journals (ARJUNA). Then, the government set some regulations to collect scientific bibliographies of Indonesian researchers. Those activities have final goal to increase the national research productivity, which is relevant in the community problem-solving. Current work in this paper also reinforced the government objective by contributing to a visual approach of science mapping for encouraging user exploration in research interest. Therefore, the next tasks for the visualizing system of SIPETRUK, another character from Indonesian folktales, are related to the recommending feature which has not been thoroughly discussed in here.

Compliance with Ethical Standard

Funding This work was funded by Institute of Research and Community Service (Lembaga Penelitian dan Pengabdian Masyarakat, LPPM) Institut Teknologi Sepuluh Nopember (ITS) Surabaya with the grant number of 1592/PKS/ITS/2019.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. C. Zhang, Z. Li, J. Zhang, A survey on visualization for scientific literature topics. J. Vis. **21**(2), 321–335 (2018)
2. A. Skupin, The world of geography: visualizing a knowledge domain with cartographic means. Proc. Natl. Acad. Sci. **101**(Supplement 1), 5274–5278 (2004)
3. J. Zhang, C. Chen, J. Li, Visualizing the intellectual structure with paper-reference matrices. IEEE Trans. Vis. Comput. Graph. **15**(6), 1153–1160 (2009)
4. K. Hu et al., Understanding the topic evolution of scientific literatures like an evolving city: using Google Word2Vec model and spatial autocorrelation analysis. Inf. Process. Manag. **56**(4), 1185–1203 (2019)
5. F. van Ham, J.J. Van Wijk, Interactive visualization of small world graphs, in *Information Visualization. 2004. INFOVIS 2004. IEEE Symposium*, pp. 199–206 (2004)
6. Z. Huang, H. Chen, F. Guo, J.J. Xu, S. Wu, W.-H. Chen, Visualizing the expertise space, in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004,* p. 9 (2004)
7. P. Federico, F. Heimerl, S. Koch, S. Miksch, A survey on visual approaches for analyzing scientific literature and patents. IEEE Trans. Vis. Comput. Graph. **23**(9), 2179–2198 (2017)
8. F. Heimerl, Q. Han, S. Koch, T. Ertl, CiteRivers: visual analytics of citation patterns. IEEE Trans. Vis. Comput. Graph. **22**(1), 190–199 (2016)
9. T. Kurosawa, Y. Takama, Predicting researchers' future activities using visualization system for co-authorship networks, in *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, vol. 1, pp. 332–339 (2011)
10. S. Latif, F. Beck, VIS author profiles: interactive descriptions of publication records combining text and visualization. IEEE Trans. Vis. Comput. Graph. **25**(1), 152–161 (2019)

11. T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, *CoRR*, vol. abs/1301.3 (2013)
12. L. van der Maaten, G. Hinton, Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)
13. T. Basu, C.A. Murthy, Towards enriching the quality of k-nearest neighbor rule for document classification. Int. J. Mach. Learn. Cybern. **5**(6), 897–905 (2014)