Academic Community Explorer (ACE) for Syntactic, Semantic and Pragmatic Document Analysis

Akansha Bhardwaj[‡§], Dominique Mercier[‡†], Hisham Hashmi[‡†], Andreas Dengel[‡†], and Sheraz Ahmed[‡]
[‡]Smart Data and Services, DFKI Kaiserslautern, Germany
[†]TU Kaiserslautern, Kaiserslautern, Germany
[§]eXascale Infolab, University of Fribourg, Switzerland
[‡]$firstname.lastname@dfki.de$, [§]$akbwaj@exascale.info$

***Abstract*—This paper presents a novel Academic Community Explorer (ACE) which performs syntactic, semantic and pragmatic document analysis of scientific publications. Firstly, ACE uses syntactic structure to extract relevant information from a scientific document. Secondly, semantic analysis is performed to derive an article based co-authorship and citation network. Finally, ACE uses these document based networks to build a complete community network for pragmatic analysis. Furthermore, scientometric analysis is performed to extract the pragmatics by analyzing authors and publication community networks through micro and macro indicators. Two novel micro indicators Senti-Index, reflecting the sentiment present in citations and, Overlap index, reflecting community behavior have been introduced. This is a step in the direction of automatic qualitative assessment of scientific documents. In addition, ACE provides a rich visualization interface which helps in exploratory analysis of the community to identify hidden patterns, e.g, isolated small groups in the community which collaborate and cite each other frequently. A feasibility study is performed on the corpus of ICDAR publications from 1993-2015 to show the insights and benefits of the ACE framework. The results reveals that ICDAR is a highly collaborative community which has most likely arrived at its 'phase transition' stage with 70% of the community closely connected to each other.**

I. Introduction

In recent years, the field of document analysis has advanced tremendously from the use of handcrafted features for layout and textual analysis [1]–[3] to deep learning [4]. Though document understanding has several applications for business and academic purposes, the application of pragmatic analysis of documents for community analysis has been overlooked.

The aim of community analysis is to explicitly focus on the quantitative and qualitative assessment of a community. This is quite important for scientific communities, as numerous scientific documents are published every year. In this context, citation analysis and quantitative metrics have been drawing interest from academia lately, attempting to replace traditional productivity indicators like h-index [5]. Although there are already some approaches available for analyzing communities by visualizing co-citation and bibliographic coupling networks [6]–[8], all of these approaches rely on meta-data of the publications, i.e., structured information like BibTex instead of the publication itself.

This paper presents a novel framework (ACE) to perform syntactic, semantic and pragmatic analysis of scientific publications/documents with an explicit aim of community analysis. Data which results this analysis can convey a lot of information about scholarly collaborations, communication, networks of scholars, and thematic trends. With the help of semantic and pragmatic analysis of ACE, different important metrics (macro and micro indicators) are computed, which help in understanding and analyzing the community from different perspectives. In addition, two novel metrics related to community behavior (Overlap Index) and sentiment of the publication (Senti-Index) are also introduced in the paper. Along with these novel features, ACE is generic and applicable to scientific publications from any community and in any format e.g., IEEE, ACM, Springer.

A. Related Works

This sections provides an overview of different approaches available for syntactic, semantic and pragmatic analysis of documents.

Syntactic document analysis includes converting raw documents to structured representation using information extraction approaches. [9], PDF box[1], pdftotxt[2] are some of the approaches to convert PDF documents into structured text. This structured representation includes title, header, keyword, abstract, and references tags as done in [10]. The next step is reference segmentation which is a challenging task owing to varying citation formats. The methods of reference segmentation can be broadly put into four major categories of template matching methods [11], [12], supervised machine learning based approaches [13]–[16], unsupervised classification approaches [17], [18], and web based look up approach [19]. Targeted data extraction has also been achieved using Named Entity Recognition (NER) approaches [20].

Semantic document analysis approaches make the data more meaningful. This includes data cleaning and transformation approaches like name resolution.

[1]https://pdfbox.apache.org/
[2]https://linux.die.net/man/1/pdftotext

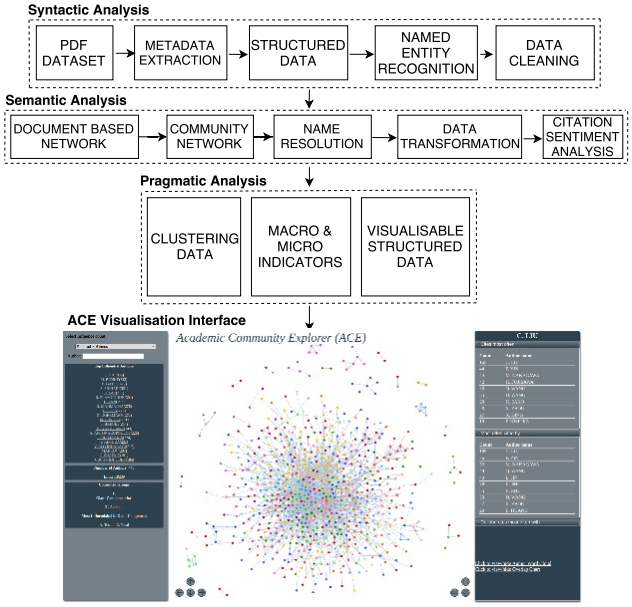


Fig. 1. Overview of the presented ACE framework.

Several approaches have been adopted in the past for the purpose of pragmatic analysis of a scientific community. [6]–[8] presented some of these existing approaches for analyzing and visualizing co-citation and bibliographic coupling networks. Some works focused on the network topology of co-authorship network of biology, physics, and smaller communities [21]–[23]. Centrality indicators [24] have been studied as a crucial determinant of the position of influence of authors in co-authorship networks. It is relevant to mention that all of these approaches rely on the meta-data of documents. None of the existing approaches use document analysis for meta-data extraction. Also, to the best of our knowledge, there has been no study on author citation networks which has been presented in this work.

II. ACE: The Presented Approach

Figure 1 provides an overview of the presented ACE framework, which is divided into three stages, *syntactic analysis*, *semantic analysis* and *pragmatic analysis*. ACE starts with syntactic analysis of document to extract important information from scientific publications. This extracted information is then used by semantic analysis to further add meaning to it. Both syntactic and semantic analysis build a baseline for pragmatic analysis, where all of the extracted information is evaluated in a global and local context of the community. The main emphasis of ACE framework is on pragmatics analysis based on author collaboration and citation networks derived from scientific publications.

A. Syntactic Document Analysis

ACE uses syntactic structure of publications to extract important information from the documents, which serves as a backbone for higher-level semantic and pragmatic analysis. An important purpose of syntactic analysis is to convert unstructured data of scientific articles to structured data. This step starts with the conversion of unstructured PDF data to text [9] followed by meta-data extraction using syntactical information. The extracted meta-data includes title, header, keywords, abstract and references. Once the meta-data is extracted, the next step is the identification of author names occurring in *header* and *references* tags from structured data extracted above using NER. Furthermore, all the sentences in text containing citation/reference to any publication are also extracted. This is very important to compute the sentiment of citation/reference.

B. Semantic Document Analysis

Semantic analysis is performed on the structured data (extracted through syntactic analysis) to add meaning to it. To convert this structured data to meaningful information, the first thing is to build a co-authorship and author citation network for each publication/document. These publication/document based networks are combined further to form a community network. Furthermore, data cleaning is performed by removing authors who never published in this conference. Another issue that plagues such a data often is of name referencing. This is due to variation in style for citation of an author. All these styles should accurately reference a common entity using the information of first name, middle name and last name provided in 'B-PER', 'I-PER' and 'E-PER' tags in Senna library [25].

In contrast to existing approaches, semantic analysis in ACE not only focuses on various networks but also emphasizes on another unique aspect i.e., sentiment analysis of citations. Here sentiment of each citation in the publication is computed based on the content of the corresponding sentence containing the citation [26].

C. Pragmatic Analysis

The purpose of pragmatic analysis is to analyze the publications in the context of whole community. To do so, first, clustering is performed on the community co-authorship network, followed by, calculation of various performance indicators for co-authorship and author-citation networks, which gives an insight of authors, publications, and whole community through different perspectives.

1) Clustering co-authorship and author-citation network: It is important to perform clustering on the community network to find important structures and patterns in it. In this context, it specifically refers to finding groups of authors who collaborate and cite each other frequently. Authors in co-authorship network are clustered using Girvan-Newman clustering approach [27], which is a hierarchical clustering approach. Here, edges are removed through an iterative process based on high betweenness centrality. The level of clustering was identified through an empirical process. Authors in author-citation network have been clustered using a graph partitioning based on *edge strength.*

2) Performance indicators: There are two kinds of performance indicators that help in analyzing a community from different dimensions, *macro indicators* and *micro indicators*. In

this work, we have explored macro indicators on co-authorship networks and author citation networks, while micro indicators have been calculated for both these networks separately.

Macro indicators in co-authorship network: Macro indicators are a series of characteristics that focus on network topology.

- Statistical summary
- Study of evolving co-authorship network
- Region based co-authorship analysis
- Publication vs. citation analysis

Micro indicators in co-authorship networks: The field of network analysis draws heavily on graphical imagery to reveal the display and patterns of links occurring within the network and uses mathematical and computational models to describe and explain those patterns. These patterns and indicators are micro-indicators which represent qualitative data such as the power, stratification, ranking, and inequality in social structures.

The co-authorship network is a special kind of social network where authors are represented as nodes and a co-authorship is denoted by the presence of an edge between two nodes. Vertex specific measures that have been explored in this work for co-authorship networks include degree centrality, closeness centrality and betweenness centrality. A high degree centrality denotes the existence of authors who collaborate very often with many other authors. These are prolific writers of a community. A high betweenness centrality is used to denote the authors who act as bridges between small sub-groups in a community and thence, help to bring the complete community together. These positions are generally occupied with head of research groups. A high closeness centrality denotes diversity of an author's domain. These authors can spread research ideas quickly.

Micro indicators in author citation networks: Author citation networks are networks, where, authors are nodes and a directed edge exists between A and B if 'A cites B' in one of its works. A basic property of an author citation network is that, when nodes are arranged according to degree centralities, a big node denotes the dominating person in a community. To the best of our knowledge, author citation networks are being explored in this work for the first time.

Few vertex specific measures explored in this work include degree centrality, indegree centrality, outdegree centrality, betweenness centrality, and eigenvector centrality. While a high indegree centrality denotes the authors who have been cited most, a high outdegree centrality denotes the authors who cite others most. Degree centrality is a combined metric for the above two centrality measures. Betweenness centrality is used to denote authors who diversely publish and communicate with others in community. These authors cite others and are cited by in a balanced way from the community. Eigenvector centrality is used to identify the authors who are most likely to receive first new research ideas.

In addition to the above-mentioned matrices, this paper also presents two novel metrics, i.e., Overlap-Index and SentiIndex.

a) Overlap-Index: An overlap index is a new metric which is being introduced in this work to study an author's diversity of influence in a community. It is a cumulative graph to glance over an author's relation with others in the community with a quantitative count of collaboration, citation and references. For an author, an overlap index shows a quantitative overlap with other authors who share a 'collaboration','cited by','cites' relation with the former. In the ACE visualization interface, we have focused on top-10 authors from each of these three categories. If the count of authors a_1,..,a_n on x-axis is more in an overlap index for author a_0 in consideration, it denotes that the author a_0 has a very diverse group with whom he/she collaborates. The count on y-axis denotes the number of times the 'collaboration', 'cited by', 'cites' relationship exists.

b) Senti-Index: Another novel metric introduced in this paper is Senti-Index [26], which expresses the total number of positive, negative, and neutral citations which an author has received in each of their articles. This is a step in the direction of automated qualitative assessment of scientific documents. Senti-index can be computed both for individual publications as well as for authors.

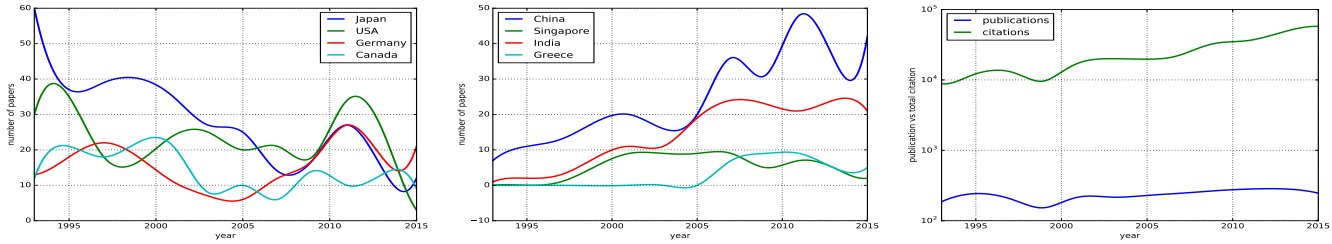
III. Analysis of ICDAR Community with ACE

To show the effectiveness of the presented framework, we present an evaluation of the ICDAR community using the ACE framework. This analysis is based on all publications in ICDAR from 1993-2015. ICDAR started in 1993 as a community on document analysis and recognition and is presently one of the main conferences in this field. At present, it has around 3500 authors participating from approximately 55 different countries.

A. Macro indicators of ICDAR community

1) Statistical summary of ICDAR: Table I provides a statistical summary of ICDAR co-authorship network. There are 3636 authors in this network, in which an average author writes 6.01 papers and collaborates with 4.75 authors. This community has a very high clustering coefficient of 0.7 which means that, there is a 70% chance of two authors being co-author if they have a mutual co-author. These numbers are relatively higher as compared to the LIS co-authorship network [22] and similar to the co-authorship networks of biology and physics constructed by Newman [23]. This number shows that ICDAR is a community where authors collaborate more frequently and widely as in the field of Biology and Physics. It is important to mention here that these results have been calculated on the PDF documents from ICDAR conference which were encoded with proper glyph to character mapping.

2) Evolution of ICDAR co-authorship network: Table II shows the evolution of ICDAR co-authorship network from 1993 to 2015. On an average, each author has more collaborators from 1993-1997 period to 1993-2015 period. This indicates that authors have collaborated more widely in recent years.



(a) (b) (c)

Fig. 2. (a) Region based authorship analysis: decreasing trend, (b)Region based authorship analysis: increasing trend (c)Publication vs. citation (shown in blue)

TABLE I
SUMMARY OF STATISTICS OF ICDAR CO-AUTHORSHIP NETWORK

Values	ICDAR	LIS[22]	Physics[23]
Number of papers	2751	10,344	98,502
Number of authors	3636	10,579	52,909
Paper per author	6.01	2.40	5.1
Author per paper	-	1.80	-
Largest component	70.84%	20.77%	85%
Clustering coefficient	0.71	0.58	0.43
Pearson Clustering coefficient	0.12	NA	0.36
Average collaborators	4.75	2.24	9.7
Average distance	5.58	9.68	5.9

TABLE II
ICDAR EVOLVING CO-AUTHORSHIP NETWORK

Year	Authors	Papers	Average coauthors	Largest Component		
				Size	Ratio (%)	Mean distance
1997	1088	634	3.23	189	17.37	5.32
2003	1859	1220	3.80	857	46.10	6.70
2009	2786	1962	4.12	1682	60.37	6.65
2015	3636	2751	4.75	2576	70.84	5.58

The value of the largest component reveals that after 1999, mean distance has decreased and the ratio of the giant component with respect to the whole community has increased from 17% in 1997 to 70% by 2015. This is similar to the work done by Barabasi [28] and suggests that ICDAR is a highly collaborative community and has probably arrived at its 'phase transition' stage where authors collaborate more frequently and widely with each other.

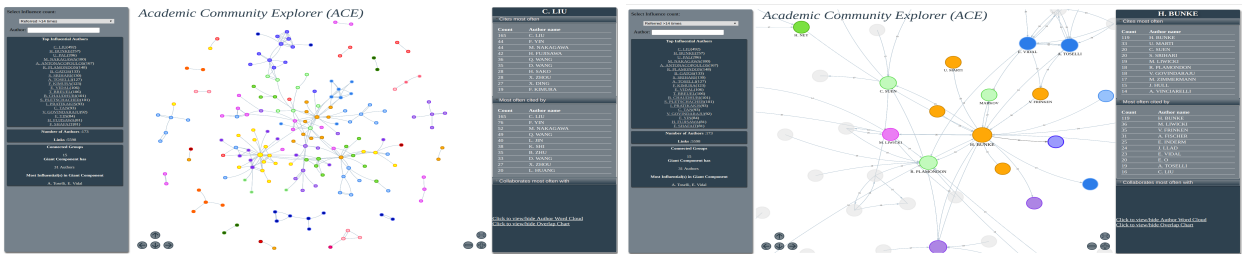
3) Region based authorship analysis: Since 1993, ICDAR community has received participation from around 55 countries. It is interesting to observe that countries like Japan, USA, Germany and Canada which have been very prolific participants in ICDAR conference in 1993-2003 make lesser contributions at present. From 2005 onwards, a peculiar trend has been observed where countries including China, India, Singapore have started participating more actively. This trend can be easily observed in Figure 2a, 2b. France has maintained its consistency and has contributed maximum publications to the community.

4) Publication vs. citation: A comparison of the number of papers and the increasing citation count has been presented in Figure 2c. It is observed that the citation tendency has increased exponentially after 2005. Though the total number of publications remain almost same, there is an 150% increase in total number of citations from 2005 to 2015.

B. Micro indicators of ICDAR community

This section provides analysis of different micro indicators (Section II-C2) in context of ICDAR community. This includes, co-author network, author citation network, word cloud and a novel measure called Overlap index and Senti-index of authors and publications.

1) Co-authorship network: ICDAR has a highly collaborative community structure where 70% of the authors are a part of the largest component.



(a) (b)

Fig. 3. (a) Academic Community Explorer(ACE) Interface, (b) Zoomed in view of ACE interface

TABLE III
NORMALIZED CENTRALITY VALUES FOR PROMINENT CONTRIBUTORS IN ICDAR

Author	Coauthorship networks			Author citation networks			
	Closeness	Betweenness	Degree	Betweenness	Eigenvector	Outdegree	Indegree
U. Pal	1.0	1.0	0.85	0.39	0.18	0.70	0.42
C. Liu	0.93	0.64	1.0	0.65	1.0	1.0	0.80
M. Liwicki	0.94	0.51	0.72	0.32	0.12	0.59	0.34
C. Tan	0.93	0.30	0.56	0.47	0.11	0.57	0.47
S. Uchida	0.92	0.51	0.74	0.24	0.11	0.44	0.17
C. Suen	0.87	0.52	0.74	0.85	0.08	0.38	0.70
D. Karatzas	0.92	0.29	0.37	0.14	0.07	0.31	0.19
K. Kise	0.85	0.17	0.45	0.15	0.04	0.22	0.20
A. Dengel	0.79	0.14	0.49	0.73	0.07	0.39	0.33
J. Llados	0.92	0.29	0.55	0.17	0.08	0.41	0.20
S. Srihari	0.83	0.27	0.32	0.38	0.01	0.17	0.69
M. Nakagawa	0.86	0.23	0.66	0.20	0.42	0.42	0.30
D. Doermann	0.88	0.21	0.28	0.65	0.07	0.35	0.57
C. Jawahar	0.65	0.21	0.18	0.12	0.06	0.27	0.15
D. Lopresti	0.78	0.06	0.18	0.24	0.03	0.20	0.22
M. Blumestein	0.88	0.14	0.26	0.06	0.05	0.27	0.11
V. Govindaraju	0.79	0.19	0.46	0.49	0.09	0.37	0.42
M. Cheriet	0.84	0.26	0.41	0.36	0.09	0.37	0.26
M. Iwamura	0.87	0.10	0.44	0.04	0.06	0.22	0.14
R. Ingold	0.84	0.22	0.46	0.22	0.06	0.36	0.16

Table III is a table of few prominent authors in ICDAR co-authorship community with respect to closeness, betweenness, and degree centrality. The values have been normalized in the range of 0-1.

2) Author citation networks: Author co-citation networks have been studied in the past and have led to conclude the presence of few major clusters or group of researchers [29]. The results from ICDAR also reflected the same presence of clusters and can be seen in Figure 3a. Nodes are colored based on the information from co-authorship network where same color represents strong co-authorship. There is a strong tendency of authors to cite their co-authors. This can be explained by the fact that generally authors choose to work closely with their group and their previous works.

The author who is cited most often in ICDAR author-citation network is C. Liu, followed by H. Bunke and B. Gatos. Head of groups generally occupy positions where they can be seen forming a bridge between different unconnected communities and holding the complete community together. H. Bunke has the highest betweenness centrality in author-citation network.

It is also possible to explore an author's relation to another represented as edge as shown in Figure 4b. Along with this, in Table III we present a list of prominent authors in author citation network with respect to betweenness centrality, eigenvector centrality, outdegree centrality and indegree centrality. In addition, ACE visualization interface offers the functionality of displaying an author's centrality, along with the sentiment of citations received by an author and the list of papers published as shown in Figure 4a.

3) Topic Cloud: Topic cloud can offer an insight into an author's domain at a glance as shown in Figure 4c. They can be used to study thematic trends in a community when represented as a time series.

4) Overlap index: An example of overlap index graph for U. Pal is shown in Figure 5 where his quantitative relation with respect to other authors in terms of 'coauthorship' , 'cited by' and 'cites' is shown.

5) Senti-Index: Senti-index in ICDAR network offers the functionality to hover over a node and view an author's positive, negative or neutral citations as shown in Figure 4a.

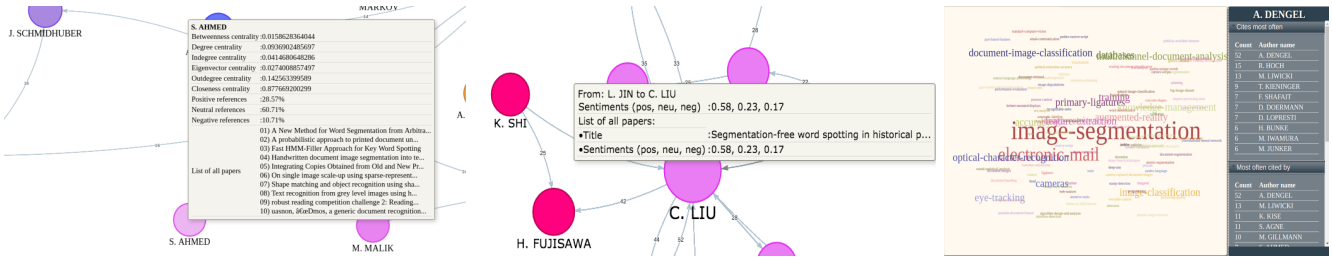


Fig. 4. (a) Node information with Senti-Index (b) Edge information with Senti-Index (c) Topic cloud for author A. Dengel

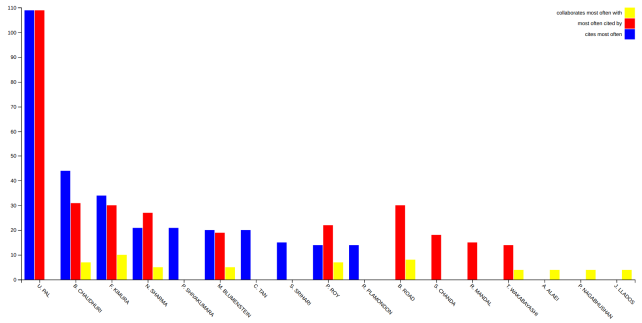


Fig. 5. Overlap index for author U. Pal

IV. Conclusion and Future Work

In this work, we have presented a framework for extracting and analyzing scholarly document meta-data for the purpose of studying a scientific community. This framework has been tested on ICDAR community from 1993-2015 and the visualization interface is live at http://www.dfki.uni-kl.de/ace/. ICDAR has the identifying characteristics of a highly collaborative scientific venue with 70% of authors being a part of the largest component. The mean distance between co-authors has considerably reduced since 1993 and the increasing ratio of the largest component indicates that ICDAR has probably arrived at the stage of phase transition. It has also been observed that countries like China and India have recently joined the community and become prolific participants. In author-citation networks, there is a clear presence of communities that collaborate and cite each other very often.

The results presented in this work are preliminary and need to be interpreted with great caution which is not possible in the narrow focus of this study. In addition to the approaches which have been presented in this work, it is important to involve additional qualitative component using interviews with some of the key members to help better understand the relationships within and between communities. In its final stage, this information can be structured and used as a recommender system for the community members to get an overall picture of their interaction to improve and reflect a better community behavior.

References

[1] S. Klink, A. Dengel, and T. Kieninger, "Document structure analysis based on layout and textual features," in *Proc. of International Workshop on Document Analysis Systems, DAS2000*, 2000, pp. 99–111.

[2] G. Nagy, "Twenty years of document image analysis in pami," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 38–62, 2000.

[3] L. O'Gorman and R. Kasturi, *Document image analysis*. IEEE Computer Society Press Los Alamitos, 1995, vol. 39.

[4] A. W. Harley, A. Ufkes, and K. G. Derpanis, "Evaluation of deep convolutional nets for document image classification and retrieval," in *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*. IEEE, 2015, pp. 991–995.

[5] J. E. Hirsch, "An index to quantify an individuals scientific research output that takes into account the effect of multiple coauthorship," *Scientometrics*, vol. 85, no. 3, pp. 741–754, 2010.

[6] N. J. van Eck and L. Waltman, "Visualizing bibliometric networks," in *Measuring scholarly impact*. Springer, 2014, pp. 285–320.

[7] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," *Journal of the Association for Information Science and Technology*, vol. 62, no. 7, pp. 1382–1402, 2011.

[8] H. Small, "Visualizing science by citation mapping," *Journal of the Association for Information Science and Technology*, vol. 50, no. 9, p. 799, 1999.

[9] J. Beel, S. Langer, M. Genzmehr, and C. Müller, "Docear's pdf inspector: title extraction from pdf files," in *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 2013, pp. 443–444.

[10] D. Bergmark and C. Lagoze, "An architecture for automatic reference linking," *Research and Advanced Technology for Digital Libraries*, pp. 115–126, 2001.

[11] I.-A. Huang, J.-M. Ho, H.-Y. Kao, and W.-C. Lin, "Extracting citation metadata from online publication lists using blast," *Advances in Knowledge Discovery and Data Mining*, pp. 539–548, 2004.

[12] G. Sautter and K. Böhm, "Improved bibliographic reference parsing based on repeated patterns," *International Journal on Digital Libraries*, vol. 14, no. 1-2, pp. 59–80, 2014.

[13] H. Han, C. L. Giles, E. Manavoglu, H. Zha, Z. Zhang, and E. A. Fox, "Automatic document metadata extraction using support vector machines," in *Digital Libraries, 2003. Proceedings. 2003 Joint Conference on*. IEEE, 2003, pp. 37–48.

[14] F. Peng and A. McCallum, "Information extraction from research papers using conditional random fields," *Information processing & management*, vol. 42, no. 4, pp. 963–979, 2006.

[15] I. G. Councill, C. L. Giles, and M.-Y. Kan, "Parscit: an open-source crf reference string parsing package." in *LREC*, vol. 2008, 2008.

[16] K. Seymore, A. McCallum, and R. Rosenfeld, "Learning hidden markov model structure for information extraction," in *AAAI-99 workshop on machine learning for information extraction*, 1999, pp. 37–42.

[17] E. Cortez, A. S. da Silva, M. A. Gonçalves, F. Mesquita, and E. S. de Moura, "Flux-cim: flexible unsupervised extraction of citation metadata," in *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 2007, pp. 215–224.

[18] D. Besagni, A. Belaïd, and N. Benet, "A segmentation method for bibliographic references by contextual tagging of fields," in *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*. IEEE, 2003, pp. 384–388.

[19] D. Huynh and W. Hua, "Self-supervised learning approach for extracting citation information on the web," *Web Technologies and Applications*, pp. 719–726, 2012.

[20] B. Powley and R. Dale, "High accuracy citation extraction and named entity recognition for a heterogeneous corpus of academic papers," in *Natural Language Processing and Knowledge Engineering, 2007. NLP-KE 2007. International Conference on*. IEEE, 2007, pp. 119–124.

[21] P. Zervas, A. Tsitmidelli, D. G. Sampson, N.-S. Chen *et al.*, "Studying research collaboration patterns via co-authorship analysis in the field of tel: the case of educational technology & society journal," *Journal of Educational Technology & Society*, vol. 17, no. 4, p. 1, 2014.

[22] E. Yan and Y. Ding, "Applying centrality measures to impact analysis: A coauthorship network analysis," *Journal of the Association for Information Science and Technology*, vol. 60, no. 10, pp. 2107–2118, 2009.

[23] M. E. Newman, "Coauthorship networks and patterns of scientific collaboration," *Proceedings of the national academy of sciences*, vol. 101, no. suppl 1, pp. 5200–5205, 2004.

[24] A. Bavelas, "Communication patterns in task-oriented groups," *The Journal of the Acoustical Society of America*, vol. 22, no. 6, pp. 725–730, 1950.

[25] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, no. Aug, pp. 2493–2537, 2011.

[26] "Senticite - an approach for publication sentiment analysis, under submission."

[27] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[28] A.-L. Barabasi, "Linked: How everything is connected to everything else and what it means," *Plume Editors*, 2002.

[29] M. Tight, "Higher education research as tribe, territory and/or community: A co-citation analysis," *Higher Education*, vol. 55, no. 5, pp. 593–605, 2008.