Science Mapping of Publications in Natural Language Processing in the Philippines: 2006 to 2020

Rachel Edita O. Roxas, Joseph Marvin R. Imperial, Angelica H. De La Cruz
National University, Manila
Manila, Philippines
{reoroxas,jrimperial,ahdelacruz}@national-u.edu.ph

Abstract

This study presents the quantitative and qualitative analyses of natural language processing (NLP) scientific literature during a 15-year period, from 2006 to 2020, of an underrepresented region, the Philippines. This aims to identify the current state of research work and further highlight research gaps for research development in the country. From the 217 Scopus-indexed publications which were extracted during June 2021, 185 publications were manually identified and screened for further analyses. The analyses reveal the collective profiles of these documents, research productivity, modalities, language polarity, human languages considered, and NLP areas considered in the studies. Conclusions and recommendations for the Philippine research community are presented.

1 Introduction

Language (otherwise known as natural language or human language) is an important means of communication, and natural language processing (NLP) allows the automatic processing of these forms of interactions. With the daily exponential increase in conversations as facilitated by the internet, the field of NLP is becoming a vital means of using computers to make sense of these exchanges. Thus, it is imperative to determine the current state of NLP research work for the continual progress of this field of study, and further propel the development of the field to the next level in the coming years. The current state of scientific literature through quantitative means such as bibliometric analyses have been conducted in the area of NLP (Lopez-Martinez and Sierra, 2020; Radev et al., 2016). These studies consistently reveal an increasing trend in NLP research within the past 20 years, highlighting the most influential authors, institutions, and countries in NLP research, and the main topics and actors in NLP research. Quantitative reviews on NLP have been wide and varied in terms of the area of investigation, as well as in terms of the NLP tasks considered. However, such reviews tend to be biased towards the main actors (such as researchers, research institutions, and funding agencies), prominent researchers and institutions, and natural languages considered in the scientific literature, which are mostly of the majority. Scientific literature from underrepresented regions that may consider indigenous languages do not tend to appear in these studies, and they tend to be ignored in these broad scoped and general studies (Bender et al., 2021). There appears to be a potential to investigate particular regions or geographic areas where there is a rich collection of human languages, which are mostly minority languages. Some bibliometric studies have been undertaken to focus on specific geographic locations. For instance, research work by (Yu et al., 2020) focused on China, while (Şahin and Candan, 2018) investigated research work within the Turk region. For NLP research literature, a bibliometric study on NLP in Russia (Barakhnin et al., 2018) looked into publication productivity and acceleration for each of the NLP areas within the specified region. One such opportunity is to consider focusing on documents of authors with Philippine affiliations and investigate the research work involving both majority

and minority languages, in the case of this study, on the Philippine languages with 186 established languages listed (Eberhard DM, 2021). Investigation on NLP work on Philippine languages and with Philippine affiliations have been considered in previous studies (Raga, 2016; Oco et al., 2016b; Oco et al., 2016a; Oco and Roxas, 2018; Roxas and Borra, 2000; Ilao J, 2012; Dita and Roxas, 2011; Oco et al., 2013; Roxas et al., 2014; Roxas and Borlongan, 2013). (Raga, 2016) reported on the progress and trends of NLP researches in the Philippines by considering the documents in 12 editions from 2004 to 2016 of one publication source, which is a local forum entitled "National NLP Research Symposium". The work focused on the various areas of NLP tasks, authorship and affiliation trends, and natural languages covered, and the resources built. Other publications in the Philippines focused on specific areas of NLP, such as machine translation (Oco and Roxas, 2018), speech language collection and processing (Ilao J, 2012; Oco et al., 2013; Roxas et al., 2014; Oco et al., 2016b), language resource collection (Dita and Roxas, 2011; Oco et al., 2016a), and on one specific domain in education (Roxas and Borlongan, 2013). Thus, to capture the unique contributions of scientific literature within one particular geographic location, the Philippines, this study aims to present a consolidated analysis of NLP work in the country. Thus, the study addressed the following research questions on NLP literature in the Scopus database within a 15-year period from 2006 to 2020 whose authors had been affiliated with Philippine institutions: 1) What is the research productivity and growth trends of these publications? 2) What are the characteristics of the publications in terms of NLP classification areas, main actors (authors, research institutions and funding institutions), natural language polarity and coverage, methodological approaches used? 3) What are the keywords that usually co-occur in these documents? 4) What are the characteristics of authors that usually collaborate and from which institutions? 5) Based on these findings, what are the NLP research gaps that local researchers and research laboratories can further nurture and address?

2 Review of Related Literature

2.1 Bibliometric Methods

Bibliometric reviews are systematic reviews that present analyses of publications on a particular area of study, over a period of time, and in a particular geographic region. These reviews based on extracted scientific literature from research databases such as Google Scholar (Martín-Martín et al., 2018), Web of Science (Mao et al., 2018), Scopus, Science Direct (Barakhnin et al., 2018), and other local sources. The analysis on the NLP work in the Philippines utilized a local venue, with the National Natural Language Processing Research Symposium proceedings as its source (Raga, 2016). Other local databases such as thesis databases and local e-libraries have been considered in other studies (Barakhnin et al., 2018). The data extraction requires a proper selection of the keywords for the database search towards the extraction of relevant documents for the study (Chen and Xiao, 2016). NLP bibliometric studies vary in their areas of focus as applied to medical research (Chen et al., 2018), and on specific NLP tasks such as sentiment analysis and opinion mining (Chen and Xie, 2020; Mäntylä et al., 2018; Keramatfar and Amirkhani, 2019; Piryani et al., 2017; Yousif et al., 2019; Zimbra et al., 2018), chatbots and conversational agents (Io and Lee, 2017), text summarization (Gupta and Gupta, 2019), expert recommendation systems (Nikzad-Khasmakhi et al., 2019) and the use of specific devices such as mobile devices (Chen et al., 2018). Visualization tools to illustrate analyses on the extracted documents include trend over time (Linnenluecke et al., 2020; Moral Muñoz et al., 2020; Hallinger and Kovačević, 2019; Lopez-Martinez and Sierra, 2020), and networks such as co-authorship and keyword co-occurrence networks using VOSviewer (Van Eck and Waltman, 2010; Van Eck et al., 2010). To add, some studies explored the use of NLP-enhanced bibliometrics (Atanassova et al., 2018), and IR-enhanced bibliometrics (Cabanac et al., 2018; Cabanac et al., 2020).

2.2 NLP in the Philippines

Qualitative reviews on NLP researches by authors with Philippine affiliations and on Philippine languages have been conducted. An initial review has been conducted by (Roxas and Borra, 2000), fol-

lowed by the work of (Dita and Roxas, 2011) but focusing on textual corpus building, both of which presented the NLP research work in the Philippines at that time. Then, (Raga, 2016) presented the trends of NLP researches in the Philippines using documents from 2004 to 2016 of the local forum entitled "National NLP Research Symposium" (NNLPRS). The work focused on the various areas of NLP that were considered in these documents, authorship and affiliation trends, and natural languages and the resource building scope. The NLP tasks considered in the study are machine translation, text summarization and generation, word sense disambiguation, information extraction, sentiment analysis, discourse analysis, language generation, named entity recognition, and textual entailment. Modes considered are texts, speech or audio, with language resource building as an important component. Over the 12-year period, the top three NLP areas of the documents considered are language resource collection, machine translation and speech processing. The study also highlighted that the affiliations that are producing the top number of documents are those authors that are coming from universities with well-established research laboratories and/or with NLP courses being offered in their universities. Likewise, these affiliations, with the most number of NLP documents are also those with more collaboration as reflected by co-authorships in the documents. English was still the predominant language of investigation in the documents studied, followed by Tagalog, Ilokano, Cebuano, Hiligaynon, Kankanaey, Kapampangan, and other Philippine languages. On the other hand, (Oco and Roxas, 2018) presented a systematic review of the NLP researches and work on one particular area of textual NLP, machine translation, within a 20-year period from 1998 to 2018, which involved Philippine languages. Aside from the documents in the NNLPRS, the review also included conference proceedings, journal issues, and works of local authors. Intellectual structure covered the approaches used, data size, and innovations applied. In addition to this work which focused on one NLP task on textual data, speech corpus building and speech processing have annually been reported at the O-COCOSDA conferences (Ilao J, 2012; Oco et al., 2013; Roxas et al., 2014; Oco et al., 2016b). A systematic review of language resource collection, annotation and modelling of Philippine languages has been presented by (Oco et al., 2016a). Aside from the studies on modals, languages, and NLP tasks, specific domains are also considered such as the work of (Roxas and Borlongan, 2013) on a systematic review of computational linguistics in the Philippines, focusing on its implications for a specific domain of usage which is the mother tongue-based multilingual education in the Philippines.

3 Methodology

This study employed a mixed method approach, considering both qualitative analysis and bibliometric analysis. The study considered publication data from the Scopus database up to 2020 of natural language processing literature of authors who were affiliated with Philippine institutions.

3.1 Conceptual Framework

This study hinges on the conceptual framework as provided by (Hallinger and Kovačević, 2019), which presents four dimensions of scientific literature mapping, namely: size, time, space and composition. First, size in this study pertains to the research productivity, impact, and co-authorship and keyword co-occurrence networks within the NLP research communities among authors who are affiliated with Philippine institutions within the given period. Second, time in this study refers to the publications based on Scopus-indexed publications up to 2020. Third, space covers the geographic distribution of Philippine affiliations of the publications being considered. It highlights the affiliation of the authors of the identified NLP publications that have been produced and indexed by Scopus. Fourth, composition refers to the intellectual structure of the publications under investigation. The intellectual structure refers to the components of these NLP publications, such as modality, language polarity, natural languages considered, and NLP tasks considered in the study. For modality, this study investigates the various modes considered in the publications such as text, speech or audio, images, videos and others.

Language polarity pertains to the number of natural (or human) languages considered in the particular document's investigation. Natural languages pertain to the specific natural or human languages that the

Yr	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	n
f	2	-	4	3	2	5	6	3	12	6	8	22	28	49	35	185
%	1	-	2	2	1	3	3	2	6	3	4	12	15	26	19	100

Table 1: Documents per year.

publication addresses and investigated. The Philippines has 186 established languages listed according to (Eberhard DM, 2021) with English, Filipino and Tagalog as the principal languages. Of the 186, 184 are living (and classified as 36 institutional, 67 developing, 36 vigorous, 34 in trouble, and 11 dying) and 2 are extinct. Thus, this study examines the publications that investigated any of these minority and underrepresented natural languages in the publication's research work. The NLP tasks cover the general natural language understanding and natural language generation, and also the specific NLP tasks (consistent with Raga, 2016) of language resource building (including text corpus building, speech corpus building, WordNet construction, ontology construction), lexical and grammatical analyses, named entity recognition, part of speech tagging, machine translation, text summarization, word sense disambiguation, information extraction, information retrieval, sentiment analysis, discourse analysis, and others.

3.2 Collection of Publication Data

For the collection of publication data, the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) procedure was adopted (Moher et al., 2009). Using the PRISMA procedure, four stages were followed, namely, identification, screening, eligibility and included. The current study was biased towards Scopus-indexed publications and those publications that are written in English. At the identification phase, the documents for further investigation were retrieved during the month of June 2021 using the Scopus database with the boolean query in the advanced search function of Scopus: ALL ("natural language processing") AND AFFILCOUNTRY ("Philippines"). This function extracted the 217 documents using the search function. This means that a main author who may had been affiliated as faculty or graduate students outside the Philippines, were not included in this study.

From the extracted data in Scopus using the previously mentioned query commands, the only metadata used for the study are: year, source, abbreviated source, affiliation, abstract, keywords, author details, publisher, conference name, and document type. One (1) non-English document which was written in Tagalog, and one (1) duplicate document were excluded from the retrieved data, resulting to 215 documents. There was no translation of the document provided. For the eligibility stage, the titles, abstracts and keywords were reviewed manually to remove the 30 non-NLP documents that are not related to NLP. There were 185 remaining documents for further analysis.

3.3 Analyses and Visualizations

Visualization tools were adopted from the graphs that were generated by the Scopus functions of Analyze-Search-Results and View-Citation-Overview, and those from Linnenluecke, et.al. (2020), Moral-Muñoz, et.al. (2020), Hallinger and Kovačević (2019), López-Herrera, et.al. (2011), Eck and Waltman (2009), and Eck, et.al. (2010), and Chen and Xie (2020). Intellectual structure (or composition) data were generated from manual tagging per document on the features identified, and counted. The VOS Viewer[1] tool was used to generate the co-authorship and keyword co-occurrence networks.

4 Results and Discussion

4.1 Publication Venues

Table 1 shows a consistently increasing trend of documents produced per year from 2006 to 2020, with the most number of documents produced in 2019 with 49 documents at 26%.

Table 2 shows the distribution of the documents by publication type. This shows that most of these

---

[1] https://www.vosviewer.com/

Doc Type	Conf Proc	Journal Art	Review
f	146	28	5
%	78.9	15.1	2.7

Table 2: Distribution of Document Type.

documents (with 146 documents at 78.9%) are published in conference proceedings.

Figure 1 illustrates the top 5 sources where the documents are published, with three (3) conference venues, one lecture series (the Lecture Notes in Computer Science which could have been produced from a conference paper presentations), and one (1) journal, with the top source being a conference proceeding. This shows a consistent trend considering all 185 documents (as shown in Table 2) that most of top sources conference proceedings. Figure 2 illustrates the CiteScore publication per year of the top 5 sources, with the top CiteScore being the Lecture Notes in Computer Science. Figure 3 presents the top 5 source citations by year. All top 5 sources have almost the same citations per year, except for the Lecture Notes in Computer Science. Figure 4 illustrates the country or territory of the affiliation of authors and/or co-authors (excluding the Philippines) with the top being the United Kingdom where possibly main authors are undertaking graduate studies or some other immersion activity in these institutions.

4.2 Funding Agencies

Figure 5 shows the top 10 funding sources which include Philippine government agencies under the Commission on Higher Education (with top 1 as CHED, and the PCARI program is also under CHED), and the Department of Science and Technology (top 5). Universities and foundations are also listed in the Top 10.

4.3 Citation Growth

Figure 6 illustrates a consistent increasing citation trend from 2006 to 2020 with an h-index of 9. This means that there are 9 documents that have been cited at least 9 times. Total citations of the 185 documents are 504, from 2006 to 2021, or an average of approximately 3 citations per document over the 16-year period.
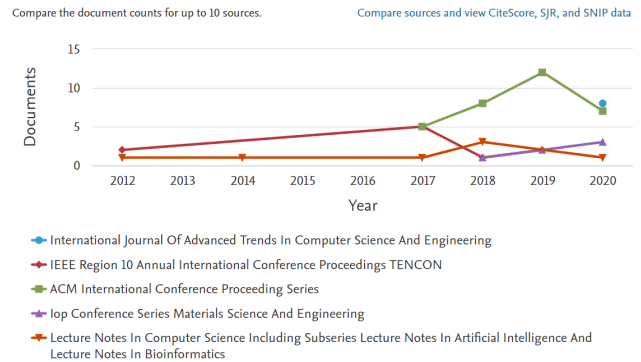

Figure 1: Documents per year by top 5 sources.
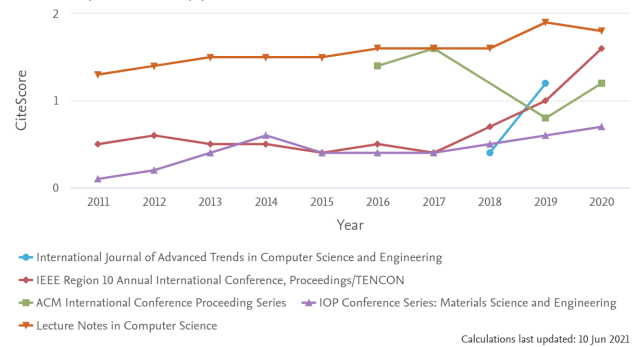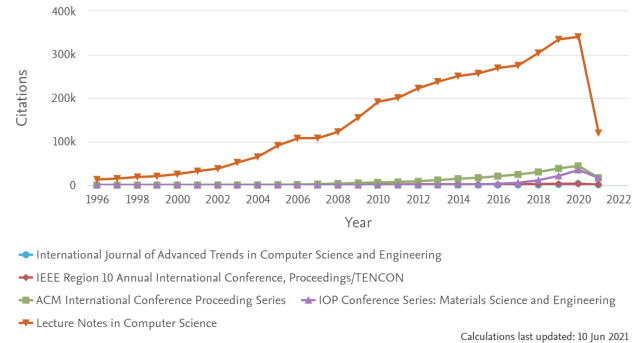

Figure 2: Top 5 sources CiteScore publication by year.

Figure 3: Top 5 source citations by year.
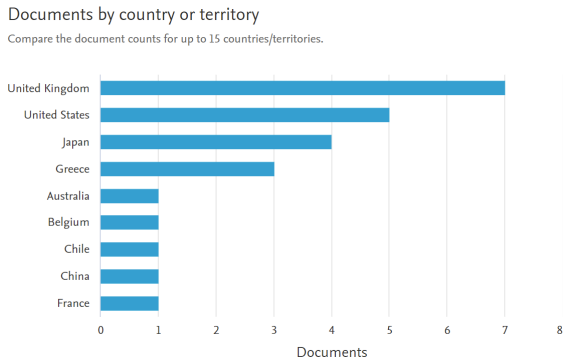
Figure 4: Documents by country or territory
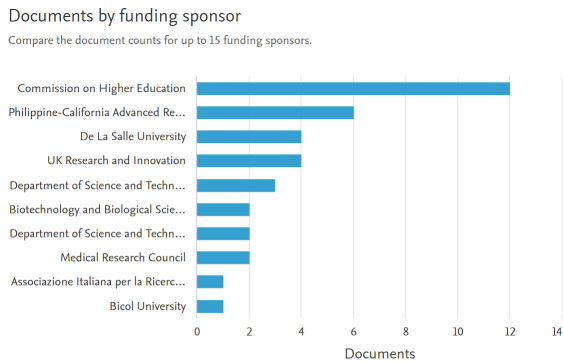

Figure 5: Top 10 funding sources.

Figure 6: Citation trend from 2006 to 2020

4.4 Intellectual Structure (Composition)

The intellectual structure (or the composition of the documents) will be described through the considered top modalities, human language polarity, human languages, and NLP areas. The top modalities (Table 3) include text processing with 151 documents at 81.6%, then followed by speech processing, and image and video processing. This shows that text processing is still the predominant mode of NLP work in the country.

Modality	Text	Speech	Image	Video
f	151	22	2	2
%	81.6	11.9	1.1	1.1

Table 3: Top modalities

Distribution of social media platforms is shown in Table 4, with Twitter leading the list, and followed by Facebook and YouTube.

Desc	SMP	Twitter	FB	YouTube
f	36	27	9	1
%	100	75	25	0.5

Table 4: Distribution of social media platforms

The distribution of top human languages that are considered in the documents (Table 5) include English with 55 documents at 30%, closely followed by the Colloquial with 50 documents at 27%, Filipino, Tagalog, Cebuano, Taglish, Ilocano, and Kankanaey. Other documents that considered one human language each are Spanish, Bikolano, Hiligaynon, and Pangasinense. Most of these languages considered are spoken by a majority number of Philippine inhabitants, except for Kankanaey. The colloquial languages refer to interaction platform in social media which include Jejemon, Filipino Textspeak or Shortcuts, Swardspeak or Gay-lingo, and Conyo. One document considered 32 Philippine languages since the study reported results of a language identification work. One document focused on the Sign Language.

Table 6 presents the distribution of the top language polarities among the documents. Most documents considered 1 language or 2 languages.

Top NLP areas considered in the documents (shown in Table 7) include classification and analy-

L	Eng	Col	Fil	Tag	Ceb	Taglish	Ilo	Kan
f	55	50	42	10	5	2	2	2
%	30	27	23	5	3	1	1	1

Table 5: Distribution of top human languages

NLP Task	CA	SA	LR	Syn	SR	MT	NIG	IE	QA	LI	C	S
f	72	25	19	14	12	10	9	4	3	3	2	1
%	39	24	10	8	7	10	5	2	2	2	1	1

Table 6: Distribution of top NLP Areas

Polarity	1	2	3	4	5
f	134	17	3	-	1
%	72.4	9.2	1.6	-	0.5

Table 7: Distribution of top language polarities

sis with 72 documents at 39%, followed by language resource building (text corpus, speech corpus, Word-Net, ontology, lexicon, presented in Table 8) with 25 documents at 24%, followed by syntactic processing, sentiment analysis, automatic speech recognition, machine translation, natural language generation, information extraction, question answering, and language identification, chatbot, and text summarization. One work focused on short hand detection from image to text. The syntactic processing documents sub-categories are distributed as shown in Table 9 among spelling checking, named-entity recognition, part of speech tagging, grammar checking, morphological analysis, and text normalization.

	Text -CB	OntoB	SpCB	LexB	Word- NetC
f	8	5	4	1	1
%	4.3	2.7	2.2	0.5	0.5

Table 8: Distribution of top language resource building documents

	Spell	NER	POST	Gram	MA	TN
f	3	3	3	2	2	1
%	1.6	1.6	1.6	1.1	1.1	0.5

Table 9: Distribution of syntactic parsing

Domains of the documents include education, disaster risk reduction, disease surveillance, cyberbul-
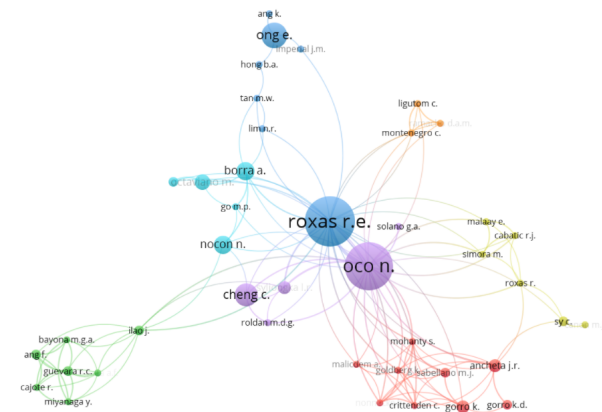

Figure 7: NLP research community via co-authorship network

lying, medicine, and business.

4.5 Research Strengths and Communities

To visualize the research communities from the examined documents, a co-authorship network (shown in Figure 7) was generated (using VOSviewer) from the author and document details of the 185 documents. The co-authorship network has circles and lines. Each circle in the co-authorship network represents an author, who had at least 2 documents in the 185 documents, and the size of the circle reflects the number of published documents made by the author. Each line between two adjacent or connected circles (or authors) represents co-authorship between the two authors, and the strength of linkages between two authors are reflected by the closeness of the distance between the two authors.

The generated co-authorship network formed 7 different research clusters, represented by the 7 colors in the network. The clusters largely represent
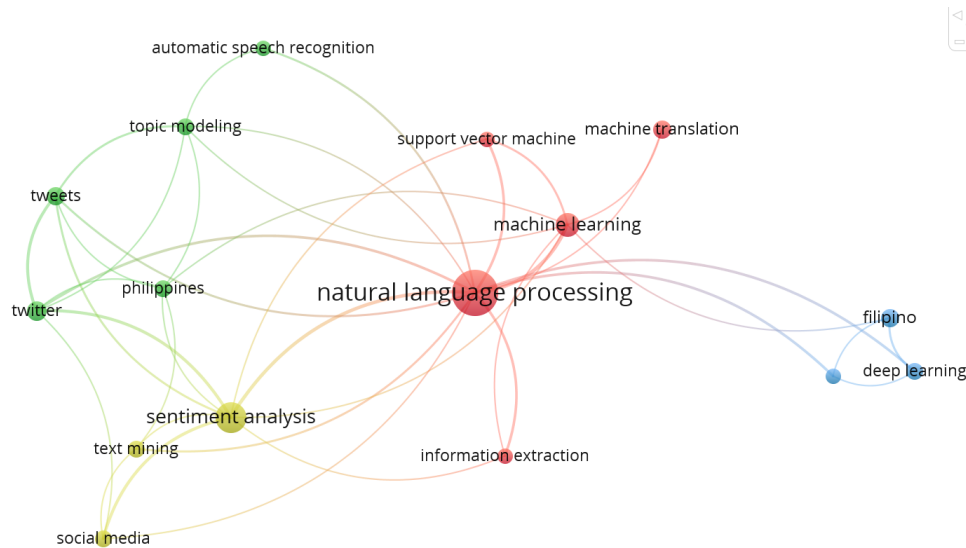
Figure 8: Prominence of research topics based on keyword co-occurrence

various institutions, and collaboration among the institutions represent the connections among the institutions. This collaboration transpired during a two-year, externally-funded research project with CHED through the Philippine California Advanced Research Institutes program as the main funding agency. The project involved the two main authors (largest circles in the middle of the network) to collaborate with more than 5 institutions (of which, 4 are Philippine institutions and 1 in California, US) where research papers are one of the outputs and were published in conferences. This shows the importance of such types of research projects as one of the main catalysts for growth and development of research productivity, as well as, an effective way to form partnerships and collaborations with other institutions not only within the country but also beyond.

To further analyze the strengths of NLP research in the Philippines, a keyword co-occurrence network was generated (shown in Figure 8) from the Scopus-generated details of the 185 documents. A minimum number of keyword co-occurrence was set to 5 to aggregate documents working on the major fields of NLP. From the results of keyword co-occurrence analysis, "natural language processing" and "sentiment analysis" have the greatest number of co-occurrences in the documents retrieved.

Sentiment analysis mostly used Twitter data due to its non-complex method of collection as there are open-source tools. These are consistent with the results shown in Tables 4 and 7. In addition, the most commonly conducted NLP tasks apart from sentiment analysis are topic modelling also using Twitter data as one of its major source, and machine translation mainly using parallel corpus from religious texts.

The enrichment of artificial intelligence approaches to NLP is also evident in the topics generated, such as machine and deep learning, and support vector machines, which are additions to the work as reported by (Raga, 2016).

5 Conclusions and Recommendations

The results of this study have shown the trends and state of NLP research work of authors with Philippine-affiliations. This study a consistently increasing trend of NLP publications produced per year from 2006 to 2020, however, very few papers of 185 were produced within the 15-year period of investigation. In terms of quality, most of these documents are conference papers rather than articles in journals, with an h-index of 9 and an average citation of 3 per paper.

Most of the documents focus on text processing using Twitter as the SM platform, and NLP tasks

mainly on classification and analysis, and sentiment analyses. Strong emphasis has been on one language per document, showing room to explore more of the 186 established Philippine languages. The results indicate possible work on other modalities, NLP tasks, and Philippine languages, as research gaps that researchers may further explore on. Two main authors collaborate with other universities and research institutions, which shows a strong collaboration network within the country. However, expansion to other Philippine universities and overseas can be explored.

This study can serve as a complement to other review documents in showing the trends of existing work, in order to guide the research network in planning for the future. It also functions as baseline reference study to monitor the trends in the coming years. It is further recommended that NLP research in the country be expanded to include the representation and reasoning using other Philippine languages, and also the consideration of the other NLP areas. It is recommended to further exploit on the use of new tools in machine learning and deep learning in the NLP research work.

References

Iana Atanassova, Marc Bertin, and Philipp Mayr. 2018. Mining scientific papers: Nlp-enhanced bibliometrics. *Frontiers in Research Metrics and Analytics*, 4:2.

VB Barakhnin, AN Duisenbayeva, O Yu Kozhemyakina, YN Yergaliyev, and RI Muhamedyev. 2018. The automatic processing of the texts in natural language. some bibliometric indicators of the current state of this research area. In *Journal of Physics: Conference Series*, volume 1117, page 012001. IOP Publishing.

Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623.

Guillaume Cabanac, Ingo Frommholz, and Philipp Mayr. 2018. Bibliometric-enhanced information retrieval: preface. *Scientometrics*, 116(2):1225–1227.

Guillaume Cabanac, Ingo Frommholz, and Philipp Mayr. 2020. Scholarly literature mining with information retrieval and natural language processing: Preface. *Scientometrics*, 125(3):2835–2840.

Guo Chen and Lu Xiao. 2016. Selecting publication keywords for domain analysis in bibliometrics: A comparison of three methods. *Journal of Informetrics*, 10(1):212–223.

Xieling Chen and Haoran Xie. 2020. A structural topic modeling-based bibliometric study of sentiment analysis literature. *Cognitive Computation*, 12(6):1097–1129.

Xieling Chen, Ruoyao Ding, Kai Xu, Shan Wang, Tianyong Hao, and Yi Zhou. 2018. A bibliometric review of natural language processing empowered mobile computing. *Wireless Communications and Mobile Computing*, 2018.

Shirley Dita and Rachel Edita Roxas. 2011. Philippine languages online corpora: Status, issues, and prospects. In *Proceedings of the 9th Workshop on Asian Language Resources*, pages 59–62.

Fennig CD. Eberhard DM, Simons GF. 2021. Ethnologue: Languages of the world. twenty-fourth edition. dallas, texas: Sil international. *Online version: http://www. ethnologue. com*.

Som Gupta and SK Gupta. 2019. Abstractive summarization: An overview of the state of the art. *Sādhanā*, 121:49–65.

Philip Hallinger and Jasna Kovačević. 2019. A bibliometric review of research on educational administration: Science mapping the literature, 1960 to 2018. *Review of Educational Research*, 89(3):335–369.

Oco N Ilao J, Roxas R. 2012. Oriental cocosda 2012 philippine country report. In *Proceedings of the Oriental-COCOSDA Conference, December 2012*. IEEE.

HN Io and CB Lee. 2017. Chatbots and conversational agents: A bibliometric analysis. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 215–219. IEEE.

Abdalsamad Keramatfar and Hossein Amirkhani. 2019. Bibliometrics of sentiment analysis literature. *Journal of Information Science*, 45(1):3–15.

Martina K Linnenluecke, Mauricio Marrone, and Abhay K Singh. 2020. Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*, 45(2):175–194.

Roberto E Lopez-Martinez and Gerardo Sierra. 2020. Natural language processing, 2000-2019—a bibliometric study. *Journal of Scientometric Research*, 9(3):310–318.

Mika V Mäntylä, Daniel Graziotin, and Miikka Kuutila. 2018. The evolution of sentiment analysis—a review of research topics, venues, and top cited papers. *Computer Science Review*, 27:16–32.

Meixin Mao, Zili Li, Zhao Zhao, and Li Zeng. 2018. Bibliometric analysis of the deep learning research status with the data from web of science. In *International Conference on Data Mining and Big Data*, pages 585–595. Springer.

Alberto Martín-Martín, Enrique Orduna-Malea, and Emilio Delgado López-Cózar. 2018. A novel method for depicting academic disciplines through google scholar citations: The case of bibliometrics. *Scientometrics*, 114(3):1251–1273.

David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and Prisma Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *PLoS medicine*, 6(7):e1000097.

José Antonio Moral Muñoz, Enrique Herrera Viedma, Antonio Santisteban Espejo, Manuel J Cobo, et al. 2020. Software tools for conducting bibliometric analysis in science: An up-to-date review.

Narjes Nikzad-Khasmakhi, MA Balafar, and M Reza Feizi-Derakhshi. 2019. The state-of-the-art in expert recommendation systems. *Engineering Applications of Artificial Intelligence*, 82:126–147.

Nathaniel Oco and Rachel Edita Roxas. 2018. A survey of machine translation work in the philippines: From 1998 to 2018. In *Proceedings of the AMTA 2018 Workshop on Technologies for MT of Low Resource Languages (LoResMT 2018)*, pages 30–36.

Nathaniel Oco, Joel Ilao, and Rachel Edita Roxas. 2013. 2013 philippine country report: Recent trends in speech and natural language processing. In *2013 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*. IEEE.

Nathaniel Oco, Leif Romeritch Syliongka, Tod Allman, and Rachel Edita Roxas. 2016a. Resources for philippine languages: Collection, annotation, and modeling. In *Proceedings of the 30th Pacific Asia Conference on Language, Information, and Computation*, pages 433–438.

Nathaniel Oco, Leif Romeritch Syliongka, and Rachel Edita Roxas. 2016b. 2016 philippine country report oriental cocosda 2016. In *2016 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*, pages 1–6. IEEE.

Rajesh Piryani, D Madhavi, and Vivek Kumar Singh. 2017. Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Information Processing & Management*, 53(1):122–150.

Dragomir R Radev, Mark Thomas Joseph, Bryan Gibson, and Pradeep Muthukrishnan. 2016. A bibliometric and network analysis of the field of computational linguistics. *Journal of the Association for Information Science and Technology*, 67(3):683–706.

R Raga. 2016. Reflections on the awareness and progress of natural language processing (nlp) research in the philippines. *Philippine Computing Journal*, 11(1):1–9.

R.E. Roxas and A. Borlongan. 2013. Computational linguistics in the philippines and its implications for mother tongue-based multilingual education in the philippines. AsiaCALL, asiacall.org.

Rachel Edita Roxas and Allan Borra. 2000. Panel: Computational linguistics research on philippine languages. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 1–2.

Rachel Edita Roxas, Nathaniel Oco, and Leif Romeritch Syliongka. 2014. 2014 philippine country report. In *2014 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*. IEEE.

Köksal Şahin and Gökçe Candan. 2018. Scientific productivity and cooperation in turkic world: a bibliometric analysis. *Scientometrics*, 115(3):1199–1229.

Nees Jan Van Eck and Ludo Waltman. 2010. Software survey: Vosviewer, a computer program for bibliometric mapping. *scientometrics*, 84(2):523–538.

Nees Jan Van Eck, Ludo Waltman, Rommert Dekker, and Jan van den Berg. 2010. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and vos. *Journal of the American Society for Information Science and Technology*, 61(12):2405–2416.

Abdallah Yousif, Zhendong Niu, John K Tarus, and Arshad Ahmad. 2019. A survey on sentiment analysis of scientific citations. *Artificial Intelligence Review*, 52(3):1805–1838.

Dejian Yu, Zeshui Xu, and Xizhao Wang. 2020. Bibliometric analysis of support vector machines research trend: a case study in china. *International Journal of Machine Learning and Cybernetics*, 11(3):715–728.

David Zimbra, Ahmed Abbasi, Daniel Zeng, and Hsinchun Chen. 2018. The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation. *ACM Transactions on Management Information Systems (TMIS)*, 9(2):1–29.