

TOWARDS BETTER BILATERAL POLICY MAKING DECISIONS: INVESTIGATION OF AUSTRALIA-RELATED ISSUES IN KOREAN MEDIA THROUGH KEYWORD ANALYSIS

¹HWA JONG KIM, ²TAE SUNG KIM, ³HYUK MIN KWON, ⁴DAE SEO PARK, ⁵ELIZAVETA SREDNIK, ^{6*}KYUNG JIN CHA

¹Professor, Kangwon National University, Department of Computing and Communications Engineering, South Korea

²Undergraduate, Kangwon National University, Department of Computing and Communications Engineering, South Korea

³Undergraduate, Kangwon National University, Department of Computing and Communications Engineering, South Korea

⁴Master, Kangwon National University, Department of Computing and Communications Engineering, South Korea

⁵PhD. Candidate, Kangwon National University, Department of Business Administration, South Korea

⁶Professor, Kangwon National University, Department of Business Administration, South Korea

E-mail: ¹hjkim3@gmail.com, ²hykwon8952@abc.com, ³fopc4131@naver.com, ⁴gentlevento@naver.com, ⁵elzbtha@gmail.com, ⁶kjcha7@gmail.com

ABSTRACT

Along with the growth of Big Data age, the value of analytics using data is improving and contributes to the new value creation. Moreover, data from news articles becomes important to identify and predict economic activities. Nowadays, a lot of studies are being done using news articles to investigate the economic phenomenon. Present study analyzes which issues have received attention in Korea by using Australian news data reported in Korea. Based on this study we aim to analyze Korea's perception of Australia and further improve relationship between Korea and Australia by helping to strengthen bilateral policy decisions by using news textual analysis.

Keywords: *Big Data, News Articles, Textual Analysis, Korea, Australia*

1. INTRODUCTION

1.1 Big Data Application

As world is changing fast with the advent of the web, Internet, and smartphones, the life-span of social networking services are also speeding up the change. According to IDC reports the amount of data generated in these environments is unimaginable and doubles every two years [1]. However, there were questions of the amount of data that can be effectively used, and of the quality and limitations of the data. This situation makes web crawling attractive. Web crawling tries to get information through the open website [2-3].

Consequently the most active place for communications such as social networking service

(SNS) sites, blogs, cafes, and news sites begin to be widely analyzed. Nowadays many businesses are using SNS semantic analyses to determine whether consumers' thoughts are positive or negative and to use them in their marketing strategies. Data that is receiving attention at the same time as SNS is news articles. They are written in a more standardized format with a header, keyword, and main text data than informal blog or SNS site. Domestic and foreign research show that when an economic news article is formatted and compiled in a chronological order, such data uses important information to identify and predict economic activity, and then is used for various analysis [4-7].

This study analyzed data on Australian and Korean news, and articles about Korea in Australia.

Although the final aim in Korea was to view a bi-directional national image of the articles related to Australia, it was to proceed in one direction only to obtain Australian data. From 1990 to 2017, we bring up Australia-related articles data and analyze the text for the selection of important words by article to identify trends in the time and trends. The primary expectation is that it could help see a shift in Korean perceptions about Australia and further help with bilateral policy decisions.

1.2 Open Data Status in Korea and Australia

The goal of open data policies is to make these public data available to anyone and to create profit. Already in developed countries such as the United States, the United Kingdom, and Japan, the democratization of public data openness, the establishment of transparent relationships with the market, and the rapid growth of open data policies are prioritized [8]. Korea and Australia have also been able to open public data via the public data portals (e.g. data.go.kr and data.gov.au) [9-11].

However, performance and influence levels remain at the low levels. In 2013, the World Wide Web Foundation announced the Open Data Barometer (ODB), which assesses the nation's open data levels. The ODB indicators were scored based on their degree of preparation, implementation and influence [12]. Korea and Australia were all rated 5th in the world (81 points) in 2017.

2. LITERATURE REVIEW

Undoubtedly great availability of information accessed through different sources such as social networking sites, blogs or news outlets enables trend prediction of various kinds that regard economical, business, political or international environment[13-15]. Scholars come across that political and economic power of a specific country such as GDP might have strong relation to news coverage [16-17].

In addition, recently various researchers emphasized how news articles can be utilized make predictions with the use of various analytical techniques. For instance, in their study Guo and Vargo on how news media in different countries influence each other in covering international news used big data analysis of 4,708 online news sources from 67 countries found that a countries political and economic status is predicted by its salience in world news [18]. From the business side news articles have been used as a primary basis for analysis in order to predict what people are thinking

about a particular product or company. In their work Goorhar and Ungar [19] processed over 100,000 news articles and blog posts for mentions of the specific products of interest and extracted phrases that were mentioned about them. Case studies presented by scholars displayed good ability to pinpoint emerging subjects in large volumes of data and highlight how they can relate to the firms interests [19]. Other researchers utilized computer-based mining techniques (e.g. Leximancer, DICTION) of news articles in supply chain industry in order to give an opportunity for business to procure their sustainable strategies [20-34]. Thus, raising importance of news text analysis in prediction and forecast making. [21-34]

3. METHODOLOGY

3.1 Data Collection

In 2016, the Korea Press Promotion Foundation has developed and disclosed a Big Kinds analysis system, which has processed natural language after accumulating news articles of 42 media organizations daily since 1990 [35]. The system enables researchers to reduce the effort of framing unstructured news data individually. Big Kinds has 25 years of news data, currently has 30 million articles stacked up and news come in real time.

Table 1: Big Kinds Extraction Data Properties and Content.

Property	Content	Property	Content
News identifier	News specific number	Case/ Event classification	Case/ Event type
Date	News article issue date	Character	News-related person
Press	Media company name	Location	News-related location
Contributor	News author name	Institution	News-related institution
Title	News title	Keyword	News-related keywords

Classifications 1,2,3	News type (politics, economy, culture, international , regional, sports, IT)	Text	News text
--------------------------	---	------	-----------

	YTN	2015. 04
Reginal News Paper	26th Press (OBS, Kyeonggi, Kangwon etc.)	1990. 11 ~

In this study, data on news articles containing Australia in the news content in the database compiled by the Korea Press Institute's Big Kinds were studied in 200,000 articles between 1990 and 2016. Information was obtained from media outlets for the Big Kinds, and the major news outlets were used (e.g. Kyunghyang Newspaper, Kookmin Daily, Munhwa Ilbo, Seoul Daily News, The World Daily News, The Seoul Daily Economic Daily, Financial Daily, and MBC Economic Daily, and Media). The file type of extracted data was an excel file and the properties and contents are organized in Table 1. Table 2 is Korea Press provided by the Big Kinds

Table 2: Big Kinds Press List

Classification	Press	Data Start Point
Major News Paper (Metropolitan)	Kyunghyang News Paper	1990. 01
	Kookmin Daily News Paper	1990. 01
	Tomorrow News Paper	2003. 01
	Everyday Economy	1995. 01
	Munhwa Daily News Paper	1996. 12
	Seoul Economy	1996. 10
	World News Paper	1990. 01
	Seoul News Paper	1990. 01
	Financial News	2001. 11
	Hankyoreh	1990. 01
	Korean Economy	1995. 01
	Herald	2000. 01
MBC	1997. 04	
SBS,	1997. 11	

3.2 Analytical Tools

To perform the analysis Python was used. Python's initial idea was designed to benefit from being easy to learn, with good program readability and ubiquitous language. It is considered to be not only a universal programming language, but also a data analysis tool that is suitable for scientific calculation and the best language for application development [36]. In addition, the absorption of the reference C++ library used for data analysis resulted in significant network effects, and the ability to express more than two dimensions of data due to the grammatically simple and effective data structures [37]. Pandas Library was used for reading and refining data. Pandas can easily process and access data using DataFrame, Series objects and, by extension, distribute, maximum and minimum calculations [38]. Jupyter Notebook was used as an analysis tool [39-41]. The open source Web application enables Jupyter Notebook to create and share documents with live codes, equations, visualizations, and text to learn the grammar of markdown. It is used primarily for data cleaning and transforming, numerical simulation and statistical modeling, and machine running, and is best suited for programming analysis with interactive computing.

3.3 Data Refining

Existing data attributes of “news identifier”, “date”, “press”, “contributor”, “title”, “integration classification 1, 2, 3”, “case/event classification”, “character”, “location”, “institution”, “keyword”, and “text” were reduced to “Date”, “Press”, “Integrated Category 1”, “Title”, “Keyword”, “Characteristic sampling”, and “text”. Attributes indicating the publication date of the article were added to the list by printing only the previous year, which led to the addition of the attribute called the year, and by printing only days to months in the same way, the attribute was added to the month. The integrated classification attributes for news type were reduced to “politics”, “economics”, “society”, “culture”, “region”, “sport”, and “IT”. Undetermined values (NaN) were classified as “Other” values.

3.4 Data Analysis

The data analysis process conducted in this study is as follows. First, the graphs and tables were used to identify the data and then proper data for analysis was selected. Then, nominal words were separated and used to extract keywords through POS tagging which was performed after morpheme analysis of selected data. POS tagging is the process of identifying a part of speech tag such as a noun or verb for an individual word [42-43]. Using extracted keywords, search the original data for news concerning the keywords was performed again to analyze the issues concerning Australia, which were focused in Korea.

Then, the number of news for each topic was identified. The number of news stories per topic is shown in Table 3, and the number of news by topic can be compared in Figure 1.

Table 3: Number of News By Subject.

Theme	Number
Others	2443
International	2023
Economy	1353
Sport	1199
Culture	974
IT	521
Politics	282
Regional	19

In the economic, international and sports issues, the number of related news items was high, and in the case of international-related news, the highest number of news articles was reported at 2023.

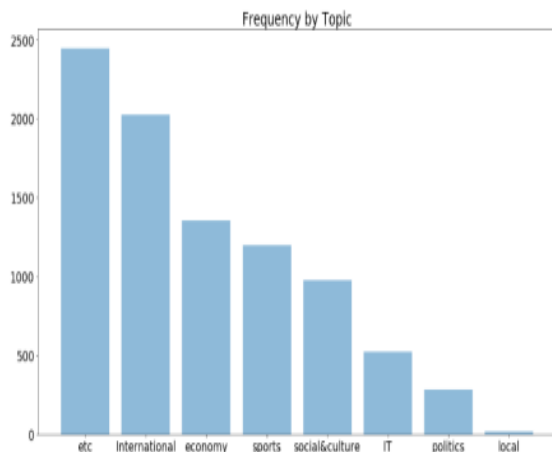


Figure 1: Thematic Frequency Graph

Figure 2 shows a lot of international, economic, sports and cultural issues being reported. That is, South Korea reports mostly on the news about Australia, including the international, economy, sports, and culture news. Figure 2 shows that out of 8814 news items, 5 percent were related to politics, society and the region, or less than 20 percent of the global news. For regional topics, the proportion is less than 0.2 percent. Therefore, it should be noted that the analysis of local keywords might not have become an issue in Korea, even if major events are analyzed through keyword analysis, since the total news frequency is low.

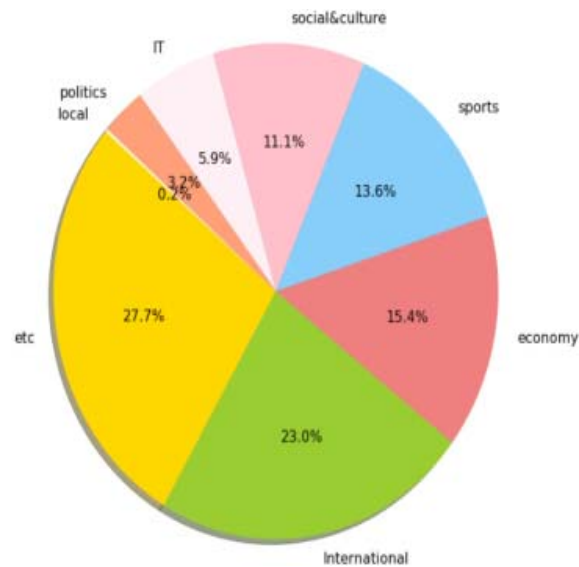


Figure 2: Percentage of Frequency Per Subject

Figures 3 show the change in the number of annual news stories between 1990 and 2017. Overall, the high number of news stories is most due to an increase in the number of media companies, so this information is not relevant to this study.

For economic properties in Figure 3, the number of news has increased in 1995, 1997, 1999, 2000, 2004, 2005, 2010, and 2014. In the case of international properties, the number of news reports in 1995, 1997, 1999, 2001, 2005, 2010, 2012, 2013, and 2015 is increasing dramatically compared to the previous years. This is due to the Korean media reporting a lot of related news on what happened in 2014, and it can be inferred that what happened in Australia was in the spotlight then in Korea. Therefore, the year when news numbers increased, the Australian issue of the year was of interest to Korea. Because of this, this study judged that the

year of a significant increase in the number of reported news compared with the previous years was a significant one and noted year for this change.

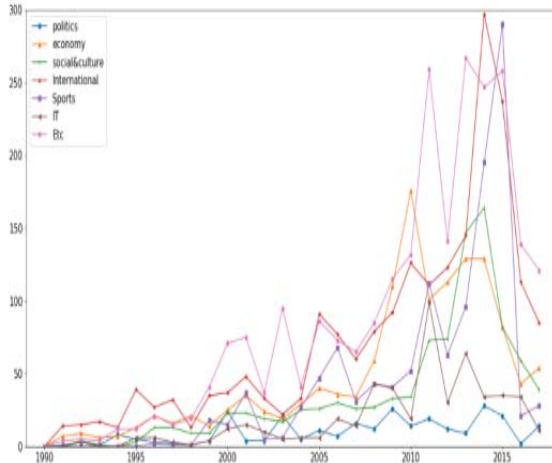


Figure 3: Number Of Stories Related To The Politics, Economy, Culture

In this study, the current year was set as the reference date to be analyzed. An analysis of news reported in Korea relative to Australia shows what happened in Australia that year that drew the attention of the Korean media, and the Australian news that was mainly reported then. Because the year of notice was defined, keywords that occurred in each year were analyzed. We used the Text Mining technique to analyze the news with KoNIPy Library [37]. KoNIPy is a Python package for Korean information processing and is available as open-source software.

Among other topics, keywords were analyzed regarding political issues first. In addition, out of the significance of political issues, 1995 was analyzed first. This study defined the 50 most frequently occurring words in each news article as keywords, that include: Australia, Japan, Canberra, Diplomacy, Wiretapping, Government, Embassy, Herald, Report, Sydney, Reuters, Union, About, Information, Morning, Chair, Wide, Spy, Activity, Sensitive, Overseas, Official residence, USA, For, China, Argument, Raise, After, Prime Minister, Country, Relation, Two countries, Dispute, Boycott, Omen, North Korea, Gangseongsan, Kitting, Last, Attention, Source, Quote, Excommunication, Agent, Infrared lines, Cutting edge, Equipment, Newspaper, Boy, Jakarta (keywords are displayed based on their ranking). At this point problem arises as it is hard to say that these words represent news as they appear very frequently, thus generic words were defined and

excluded from this study, by utilizing code displayed in the Figure 4.

The final keywords with the above code minus general words are as follows (in ranking order): Diplomatic, Wiretapping, Embassy, Reporting, Reuters, Association, Information, Morning, Wide, Espionage, Activities, Trade, Overseas, Embassy, Argument, Reporting, Country, Relationship, Dispute, Intelligence, Attention, Controversy, Attention, Gangseongsan, Normalization, Content, Message, Meantime, Cooling off, Change, Interest, Periodical.

```

from collections import Counter

def common_word_by_year(temp_data,
generalword, n):

    words = []

    for each in temp_data['text']:
        for x in tokenize(each):
            words.append(x)

    tokens = []

    for word in words: if len(word) > 1 and word
not in generalword:
        tokens.append(word)

    count = Counter(tokens)

    return pd.DataFrame(count.most_common(n))

temp_data =
news_politics[news_politics['year']==1995]
df = common_word_by_year(temp_data,
generalword, 50)

```

Figure 4: Codes For Generic Words Exclusion

According to these keywords, the 1995 Australian incident is likely to have been confirmed by the Australian government's extensive espionage activities against the Japanese Embassy in Canberra, collecting sensitive trade and diplomatic information from the Sydney Herald. Data indicate that out of the five political news collected in 1995, news covering the event accounts for the largest percentage of four. That is, among political issues in 1995, the most interesting case in Korea was the Sydney Morning Herald newspaper report on Australian spies. In the same way, international news can be analyzed. The significance of international-themed news is in 1995, 1997, 1999, 2001, 2005, 2007, 2010, 2012, 2013, and 2014. The

analysis below was conducted with the base year set at 2014, when the highest number of news stories was reported.

International keywords for the year of 2014 included provocative words such as Café, Arming, Islam, and Control. According to these keywords major events happened that shocked nations such as incident in Rinz café that occurred on December 15, 2014, when Muslim extremist held more than 20 hostages including employees and guests inside.

We extracted the monthly top keywords for 2014, in which the distribution of news articles was greatly increased, and identified major concerns of each month. Table 4 is a summary of primary keywords only, so we do not know the details of the article, but it provides a rough overview of the major events in Australia in 2014 through our keywords.

Table 4: Top 10 Monthly Keywords in 2014.

January	'Australia open', 'Competition', 'Melbourne', 'Single', 'World', 'Tennis', 'Sydney', 'Park', 'Lina', 'Ranking'
February	'Korea', 'Best', 'Victory', 'America', 'Pure Silver', 'Tour', 'Local', 'Sydney', 'Low-under', 'Time'
March	'Running Man', 'Ryu Hyun-Jin', 'Kim Woo-Bin', 'Korea', 'Crocodile', 'Sydney', 'Dodgers', 'Kangaroo', 'Giant', 'Member'
April	'Prime Minister', 'Korea', 'Abbott', 'Search', 'Free trade', 'Agreement', 'Formula', 'Ordinary', 'Minister', 'Tony'
May	'Imdami', 'X Factor', 'Program', 'Audition', 'Victory', 'Cookie', 'Local', 'Popular', 'Morning', 'Seoul'
June	'Spain', 'Brazil', 'Chilly', 'Group', 'League', 'Korea', 'Netherlands', 'Games', 'Time', 'Stadium'
July	'Crush', 'Tourist Bus', 'Tourist', 'Korean', 'Bus', 'Grey', 'Ocean', 'Rod', 'Burn', 'Cliff'
August	'Sydney', 'Branch Office', 'Dictionary', 'Account', 'Prime Minister', 'Foreign Exchange', 'Local', 'Service', 'Rad', 'Issuance'

September	'Korea', 'Australian Dollar', 'Sightseeing', 'Sydney', 'Camping', 'Cooking', 'Armature', 'Time', 'Exchange Rate', 'Local'
October	'Government', 'Iraq', 'World', 'Most', 'Operations', 'Air Strike', 'Japan', 'Local', 'America', 'Time'
November	'Meeting', 'Normal', 'President', 'Korea', 'China', 'Brisbane', 'Japan', 'Agreement', 'Sydney', 'Free Trade'
December	'Hostage', 'Sydney', 'Local', 'Cafe', 'Time', 'Korea', 'Armed', 'End', 'Police', 'Morning'

Critical issues were found, for instance in January, the tennis tournament took place which was identified with the keywords of Australian Open, tennis and singles; in March, baseball games were played, which was understood through keywords of Ryu Hyun Jin and the Dodgers; in July, there were bus crashes, which was pointed out keywords such as crashes, tourist buses and cliffs. In addition, we analyzed the number of news articles covering the keyword by month in order to identify whether the above keyword is an event that can represent the month. First, the number of news articles per month was 70, 77, 248, 101, 42, 113, 87, 29, 23, 44, 85, and 177 from January until December. The greatest number was posted in March and the lowest number was posted in September.

As it is shown in Table 5 and 6, the extracted keywords were distributed to a large number of articles on monthly basis, and it was confirmed that the keyword extraction through TF-IDF is capable of identifying issues related to Australia [44-46]. In September, it was possible to deduce that there were no special issues even though there were few articles, but it was clear that the September keyword was a keyword frequently appearing in other months, making it more clear that there was no special issue.

On the other hand, in August, although the frequency of news articles is low, bank-related keywords such as 'branch office', 'foreign exchange bank', 'service', 'issuance', and 'account' have confirmed that banking articles were covered.

Table 5: Number Of Articles Per Keyword By Month In 2014. (January to June)

Keyword/Month						
	Jan	Feb	Mar	Apr	May	Jun
a	36	1	0	0	0	0
b	6	18	11	1	14	2
c	0	3	56	0	0	0
d	0	5	4	32	1	7
e	0	0	0	1	14	0
f	7	0	0	0	3	40
g	2	0	0	2	0	0
h	0	0	5	2	0	0
i	15	16	4	3	5	2
j	0	0	0	0	0	1
k	0	8	3	4	4	3
l	0	0	0	0	0	0

Note: a – Australia Open, b – Championship, c – Running Man, d – Free trade, e – Imdami, f – Spain, g – Crush, h – Point, i – Australian Dollar, j – Iraq, k – Meeting, l – Hostage.

As shown in Figure 5, there are keywords that tend to be high in amount only in the January, February, March, April, May, June, July, November and December. Major keywords extracted during these months were generally pointing directly to one or two issues. For example, in January, the tennis competition keyword "Australian Open" showed a high frequency, and other keywords in January were also relevant to related keywords. On the contrary, the major keywords in September are not particularly high. As shown in Figure. 5, 'Australian dollar', which is the keyword of September, was more likely to appear in January and February as well.

Major international events by year are as follows: Friction between France and Australia over the resumption of nuclear tests (1995), News of Korean gangs in Australia (1997), East Timor Dispute (1999), Australian Afghanistan Refugee Treatment (2001), News on Prince Charles (2001), Australia Strasfield comfort women's prize installation (2007), Australian Central Bank interest rate news (2010), Gilad Related News (2012), Murder case in Brisbane Warhol (2013), Sydney hostage (2014).

Table 6: Number Of Articles Per Keyword By Month In 2014. (August to December)

Keyword/Month						
	Aug	Sep	Sep	Oct	Nov	Dec
a	0	0	0	1	2	0
b	0	1	3	1	5	4
c	0	0	0	0	0	0
d	0	0	2	4	26	20
e	0	1	0	0	0	0
f	0	1	0	0	1	0
g	35	0	0	0	0	2
h	0	5	0	0	0	0
i	6	0	3	2	7	6
j	1	0	1	10	0	2
k	0	0	3	6	26	2
l	0	0	0	0	0	105

In addition to identifying major issues yearly or monthly, we tried to understand the correlation between the number of Korean tourists and the number of news articles in Korea. In order to select news articles related to travel or tourism, total of about 200,000 Australian news articles were examined in the Word2Vec model [47-48], and then keywords related to 'travel', 'tourism', 'events' were extracted. The Word2Vec model is a model that examines the whole document set and learns words related to specific words or words related to them and then further extracts them. The related word is numerically expressed as a value between 0 and 1, and the closer it is to 1, the closer the word to the word.

After that, news articles including keywords related to travel and tourism events extracted from Word2Vec model were selected.

The number of tourists and news stories was relatively low compared to articles published in 2009 ~ 2016.

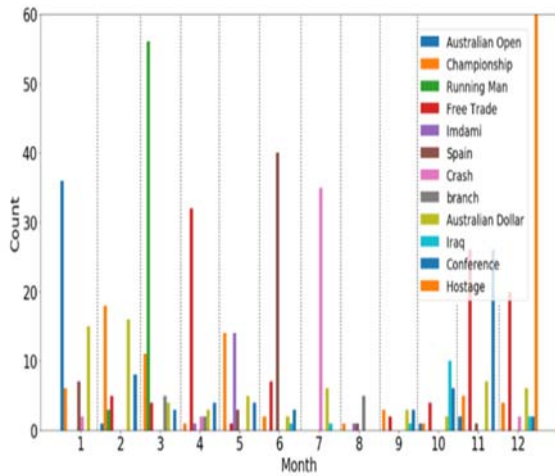


Figure 5: Graph Of The number Of Articles Per Keyword By Every Month Of 2014

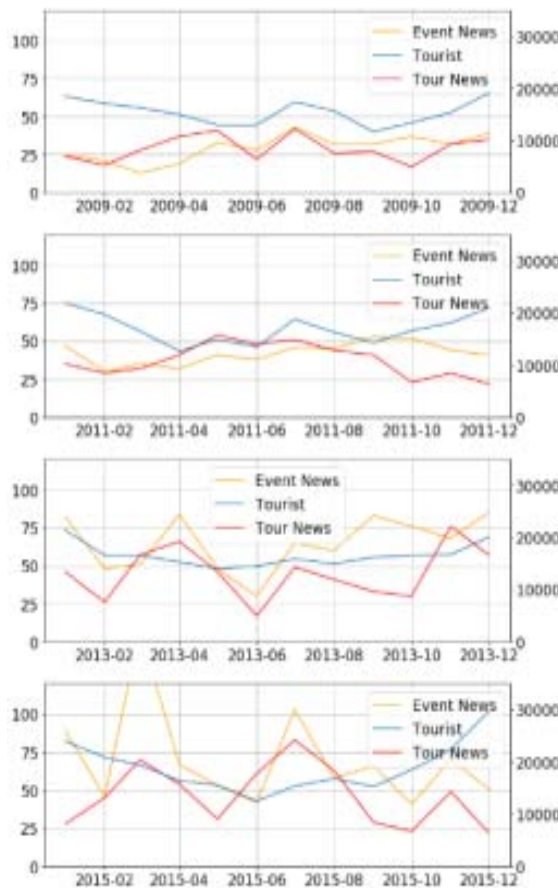


Figure 6: Comparative Graphs Of Events (Orange), Tourism News (Red), Tourist (Blue) Numbers (2009-2015)

In Figures 6 and 7, the left side shows the number of articles and the right side shows the graph of the number of tourists on monthly basis. Figure 6 shows that the number of Korean tourists

in Australia was the highest in January and December, and decreased in the months between them. Events, and tourism news show similar trends, but it is hard to find any relation with the number of Australian tourists.

However, repeated patterns were found between October and December when the number of tourists increased. The number of events and tourism news increased in November and then further decreased in December. This was due to the increase in the number of tourists due to Christmas and Boxing Day events in Australia in December. As a result, there was an increase of the news articles in October and November that covered Australian events and trips in December

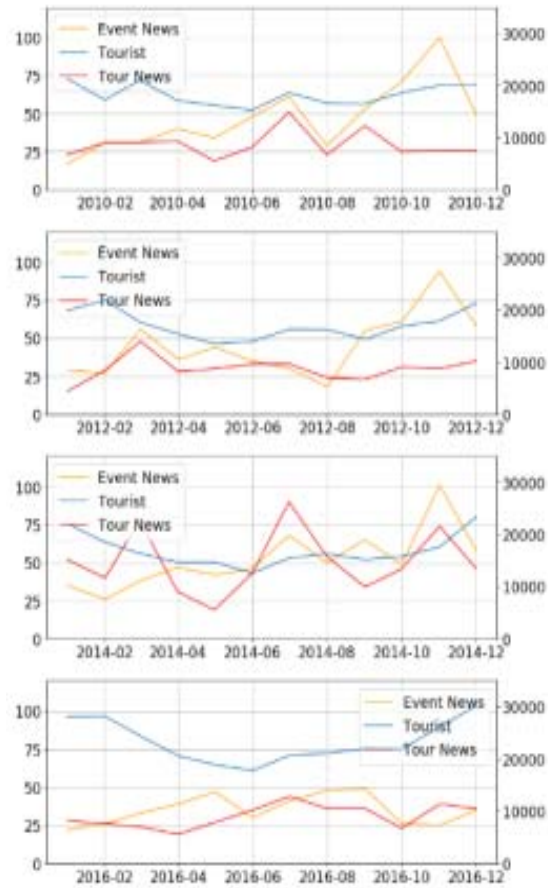


Figure 7: Comparative Graphs Of Events (Orange), Tourism News (Red), Tourist (Blue) Numbers (2010-2016)

Although local press was lacking, Australian news articles were constantly posted. Thus, issues that occurred during a specific period could be identified through keyword analysis. In addition, an analysis to identify issue keywords in news articles that determine the number of Australian tourists was conducted through events and tourism news, however it was difficult to

identify issues except for the period between October and December.

4. CONCLUSION

In this study, keywords were analyzed through speech tagging techniques and the extraction of the poorest words, and the issues that were reported in Korea of interest during Australian related events were reviewed. As a result of analysis of social news, the total number of articles was very small, so the total number of cases related to the case was not very high in Korea, despite finding out the relevant issues through the most influential keywords.

Meanwhile, this study analyzed the Australian cases that have become an issue in Korea by setting Korean news as the analysis data set. In Australia, public data was obtained from data, but the data was difficult to obtain due to a lack of search engines for media companies. If Australian news data are obtained later, it is possible to estimate how the two countries may have an awareness of the issues related to Korea, which have become an issue in Australia using similar analysis methods.

If news analysis by Korea and Australia is added, it can analyze what Korea and Australia were interested in each other, not the cases that were of interest to the entire international community.

We can also understand the issue of how Korea's perception of Australia and Australia's perception of Korea, changes over time if emotional analysis is added as part of the study series. When analyzing keywords as a limitation of this study, the most frequent words of the entire article were not determined to be general word, and therefore, the analysis cannot be performed automatically.

In addition, the possibility of analyst's subjectivity is high and the result lacks objectivity. To prevent this, the analysis was carried out using the tf-idf method frequently used in information search theory, but if the analysis was not carried out continuously in a single article, the articles on keywords were analyzed in the most influential case [44-46]. More reliable results could be obtained by improving the tf-idf method or by using other effective keyword extraction algorithms [49]. Frequently seen events have a strong influence on people's perceptions that remain long in their heads. Events in other countries reported on the news account for a large portion of people's image of the country, especially since there is less

exposure in cases abroad. From this analysis, we can estimate which Australian cases Koreans were interested in and would remember. A bilateral policy based on this understanding would lead to mutual understanding and trust.

As the role of the Asia-Pacific region in global politics and economics grows and thus cooperation between regional countries increases, this study is an important indicator and will serve as an essential indicator when determining the ultimate direction of diplomatic or bilateral policies.

ACKNOWLEDGEMENTS: This study was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No: 2017-0-00333, User-Friendly UI/UX Technology for Intelligent Devices) and Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No: 2018-0-00261, GDPR Compliant Personally Identifiable Information Management Technology for IoT Environment)

REFERENCES:

- [1] J. Y. Lee, "Data Big Bang, Big Data (Trend of BIG DATA)", *Broadcasting Communications Journal*, 2012,
- [2] Chandni Saini., Vinay Arora., "Information retrieval in web crawling: A survey", *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2016, pp. 2635-2643
- [3] Christopher Olston, Marc Najork, "Web Crawling", *Foundations and Trends in Information Retrieval*, Vol 4, No 3, 2010, pp. 175-246
- [4] Newman, D., Chemudugunta, C., Smyth, P., & Steyvers, M, "Analyzing entities and topics in news articles using statistical topic models.", *In Intelligence and Security Informatics*, 2006, pp. 93-104, Springer Berlin Heidelberg
- [5] S. J. Lee, H. J. Kim, "Keyword Extraction Method in Electronic News Using Modification of TF-IDF", *Korean Journal of Electronic Commerce*, Vol. 14, No. 4, 2009, pp. 59-73.
- [6] A. J. Young, A. K. Bin, S. Min, "Analysis of Ebola Topics by Media Using Text Mining.", *Journal of the Korean Society for Library and Information Science*, Vol. 50, No.2, 2016, pp.289-307.
- [7] T. C. Jo, "News Article Classification based on Representative Keywords in Categories",

- pp194-198, *The Proceedings of CIMCA 99: Intelligent Image Processing, Data Analysis & Information Retrieval*, 1999.
- [8] H. M. Chung, "Government 3.0, open data from an IT perspective, and its implications. Korean Association of Public Administration", 2013, pp. 337-337
- [9] S. A. Kwon, S. H. Beak, J. Y. Lee, "Current Domestic Public Open Data Status and Case Analysis", *Korean Society for Information Society*, 2015, pp. 37-45
- [10] S. H. Song, J. Y. Kim, J. Y. Lee, "A Study on Public Open Data Status and Activation Plan for Use", *Korean Institute of Communications and Information Sciences*, 2013, pp. 854-855
- [11] Ubaldi, B, "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", *OECD Working Papers on Public Governance*, No. 22, 2013
- [12] World Wide Web Foundation, "Open Data Barometer", 2013
- [13] Shamanth Kumar, Fred Morstatter, Huan Liu, "Twitter Data Analytics", Springer, 2013
- [14] Shirley Ann Williams, Melissa Terras, Claire Warwick, "What people study when they study Twitter: Classifying Twitter related academic papers", *Journal of Documentation*, Vol. 64, No. 3, 2013, pp. 1-74
- [15] Alvaro. Ortigosa, José. M.Martín, Rosa. M.Carro, "Sentiment analysis in Facebook and its application to e-learning", *Computers in Human Behavior*, Vol. 31, 2014, pp. 527-541
- [16] Segev, E. "International news online: Global views with local perspectives.", Boston, MA: Peter Lang. 2016
- [17] Golan, G. J. "Where in the world is Africa? Predicting coverage of Africa by US television networks." *International Communication Gazette*, Vol. 70, No. 1, 2008, pp. 41-57
- [18] Guo, L., & Vargo, C. J. "Global Intermedia Agenda Setting: A Big Data Analysis of International News Flow.", *Journal of Communication*, Vol. 67, No. 4, 2017, pp.499-520
- [19] Goorha S., Ungar L. "Discovery of significant emerging trends." In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. Washington, DC, USA, 2010, pp. 57-64
- [20] D. Kim, S. Kim, "Sustainable supply chain based on news articles and sustainability reports: Text mining with Leximancer and DICTION", *Sustainability*, Vol. 9, No. 6, 2017, pp. 1008
- [21] M. S. Kim, H. M. Um, "The Study on Recent Research Trend in Korean Tourism Using Keyword Network Analysis", *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 17, No. 9, 2016, pp. 68-73
- [22] Y. J. Lee, J. H. Yoon, "A Study on Utilizing SNS Big Data in the Tourism Studies", *International Journal of Tourism and Hospitality Research*, Vol. 28, No. 3, 2014, pp. 5-14
- [23] M. H. Kang, H. G. Son, "A Study on Trends of Rural Tourism in 2017-2018 through Keyword Analysis", *Korean Journal Community Living Science*, 2018, pp. 82-82
- [24] H. C. Kim, M. J. Kim, H. J. Shin, "Method for analysis of social data service design and implementation of Jeju tourism trends", *The e-Business Studies*, Vol.15, No.3, 2014, pp. 173-193
- [25] S. J. Lee, Y. N. Chun, "Examining Public Opinion on Tourism Using Social Media Analytics", *GRI REVIEW*, Vol. 18, No. 1, 2016 PP. 182-109
- [26] Mirjana Pejic Bach, Zrinka Marušić, Markus Schatten, "Data Mining Applications in Tourism: A Keyword Analysis", *Central European Conference on Information and Intelligent Systems*, 2013, pp. 26-32
- [27] T. H. Kim, H. S. Kim, "A Study on the Semantic Network Analysis of Food Tourism Using Big Data", *Journal of Leisure Research*, 2017, pp. 372-379
- [28] J. S. Kim, "A Study on the Invigoration of Sport Tourism Using Big Data", *Journal of the Korea Entertainment Industry Association(JKEIA)*, Vol. 11, No. 3, 2017, pp. 99-109
- [29] S. Y. Ryu, S. W. Yoo, "A study on the perceptions and interests of tourists to Gangwon-do - Focusing on social media big data analysis", *International Journal of Tourism and Hospitality Research*, Vol. 31, No. 2, 2017, pp. 63-81
- [30] T. H. Kim, J. W. Byun, "A Study on Tourist Resource Activation Through The Analysis of Press Reports", *Academy of Global Hospitality and Tourism Journal (AGHTJ)*, Vol. 26, No. 2, 2017, pp. 19-33
- [31] H. B. Nam, S. W. Jeon, Y. B. An, D. S. Yu, B. H. Park, "Analysis of Tourism Research Trend using Text Mining", *The Korea Academic Society Of Tourism And Leisure*, 2017, pp. 283-288
- [32] S. S. Oh, S. H. Kim, "A Study on the Effect of SNS Tourism Information Characteristics towards the Choice of Tourist Destination",

- Journal of Industrial Economics and Business*, Vol. 25, No. 5, 2012, pp. 3255-3271
- [33] M. J. Jung, H. L. Rho, M. S. Park, "Reaction survey of tourist who railway utilizing SNS big data", *The Korean Society For Railway*, 2018, pp. 456-461
- [34] J. Y. Park, G. Kim, C. Y. Kim, H. J. Oh, "A Study on Tourism Resource Strategy of Film Location using Social Bigdata based on SNS Trend Analysis of Jeonju Area", *The Journal of the Korea Contents Association*, Vol. 16, No. 11, 2016, pp. 477-487
- [35] E. B. Lee, J. O. Jang, J. S. Baek, "A Study of Multicultural Space in Seoul : Analysing the Coverage of Foreign Communities with News Big Data Analytics 'BigKinds' for 27 Years", *Journal of media economics & culture*, Vol.15, No. 2, 2017, pp. 7-43.
- [36] Wes McKinney., "Pyhton for Data Analysis", *O'Reilly*, 2012
- [37] E. .L. Park, S. Cho, "KoNLPy:Korea natural language processing in Python", *26th Annual Conference on Korean Information Processing*, 2014.
- [38] Fabio Nelli., "Python Data Analytics", *Apress*, 2015
- [39] Bernadette M. Randles, Milena S. Golshan, Irene V. Pasquetto , Christine L. Borgman, "Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study", 2018
- [40] T. Kluyver et al., "Jupyter notebooks—A publishing format for reproducible computational workflows", in *Proc. ELPUB*, 2016, pp. 87–90.
- [41] Pérez. F., Granger. B.E. "IPython: A System for Interactive Scientific Computing", *Computing in Science Engineering*, Vol, 9, No. 3, 2007, pp. 21–29
- [42] Chopra D., Itmatour N. J., "Natural language processing with Python", *ACORN*, 2017
- [43] G. I. Lee, E. H. Lee, J. H. Lee, "Sequence-to-sequence based Morphological Analysis and Part-Of-Speech Tagging for Korean Language with Convolutional Features", *Journal of KIISE*, Vol. 44, No. 1, 2017, pp.57-62
- [44] E. S. You, G. H. Choi, S. H. Ki, "Study on Extraction of Keywords Using TF-IDF and Text Structure of Novels", *Journal of The Korea Society of Computer and Information*, Vol. 20, No. 2, 2015, pp. 121-129
- [45] D. S. Park, H. J. Kim, "A Proposal of Join Vector for Semantic Factor Reflection in TF-IDF Based Keyword Extraction", *Journal of KII*, Vol. 16, No. 2, 2018, pp. 1-16
- [46] Juan Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries ", *In First International Conference on Machine Learning*, New Brunswick:NJ, USA, 2003, Rutgers University
- [47] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean., "Distributed representations of words and phrases and their compositionality", *In Advances on Neural Information Processing System*, pp. 3111–3119
- [48] Le. Q.V., Mikolov. T., "Distributed representations of sentences and documents", *Proceedings of the 31st International Conference on Machine Learning*, Vol 32, 2014
- [49] S. J. Lee, H. J. Kim, "Keyword Extraction from News Corpus using Modified TF-IDF", *Journal of Society for e-Business Studies*, Vol. 14, No. 4, 2009, pp. 59-73