Media Analysis of Digital Literacy Trainings for Older adults in South Korea

한국의 노인 대상 디지털 리터러시 훈련 사례에 관한 언론 보도 분석

Hyeon , Pyo

School of Computing, Korea Advanced Institute of Science, megatwins@kaist.ac.kr

Keywords and Phrases: Digital Literacy, Digital Divide, Older adults, Media Analysis, NLP

1 INTRODUCTION

Training digital literacy is a promising approach to bridge the second-level digital divide. Of various demographic features, age is one of the most important factors in the problem of digital divide. Therefore, it is important both private and public sectors to contribute to plan, conduct and support digital literacy trainings targeting older adults. Qualified database of previous training sessions and following analysis is required to mine useful insights by reviewing the plans and sessions in various aspects. Therefore, this paper aims to conduct a media analysis of digital literacy trainings scoped in older adults in South Korea, construct its database, and add some of macroscopic analysis to showcase how this media analysis and database can be used to identify trends and insights for digital literacy trainings in future.

2 BACKGROUND

2.1 Digital Divide and Digital Literacy

Digital divide refers the gap between individuals, households, businesses and geographic areas at different socio-economic levels with regard both to their opportunities to access information and communication technologies (ICTs) and to their use of the Internet for a wide variety of activities [1] (OECD, 2001). The digital divide has been described as a critical topic for social justice in the twenty-first century [2] (Rogers, 2016).

Digital divide is conceptualized into two levels. First-level digital divide refers to the binary division between people who had access to computers and the Internet and those who did not have that access [3, 4, 5] (Lythreatis et al., 2022; Dewan and Riggins, 2005; Hoffman et al., 2000). Second-level digital divide describes gap in digital skills [6] (Hargittai, 2002) and digital usage [7] (Van Dijk, 2005). This problem is also termed ‘digital inequality’ to highlight the aspect of social injustice and has been receiving a lot of attention by researchers [6] (Hargittai, 2002).

The most straightforward approach to bridge second-level digital divide is to raise minorities’ digital skills. A group of researches focused on analysis of digital skills necessary to bridge the divide [8, 9, 10] (Scheerder et al., 2017; Mossberger et al., 2003, Van Deursen and Van Dijk, 2011). Digital literacy is a new term emerged to describe such skills, and is defined as a mindset that enables users to perform tasks in digital environments and to both easily and effectively access the wide range of knowledge embedded in the digital environment [11, 12, 13](Blažič et tal., 2020; Martin 2008; Van Laar et al. 2017).

2.2 Digital Literacy of Older Adults in South Korea

This paper reviews previous approaches focused on improvement of older adults’ digital literacy in South Korea. Age is one of the demographic factors that is most frequently linked to digital divide [3, 14] (Lythreatis et al. 2022; United Nations 2012) and shortcomings in older adults’ digital literacy is a major source of the age-related digital divide.

In case of South Korea, it was found that the higher the digital literacy of older adult, the higher the social satisfaction [15] (김학실, 2020). However, level of older adults' digital literacy is very low in the nation. Level of digital informatization of the elderly aged 55 years or older is only 63.1% of the general public [16](과학기술정보통신부, 한국정보화진흥원 2018). Usage rate of mobile banking and other mobile applications is less than 1%. Researches imply urgent need for policies to improve the older adults' digital literacy of the older adults and shift the society toward digital inclusion [15] (김학실, 2020).

There's no previous review nor database of digital literacy improvement programs targeting older adults that covers various sectors of government, local government, civil society, and corporations in South Korea. 박하나 et al. have reviewed the educational courses on digital literacy in 5 countries; South Korea, UK, Australia, Ireland, and Canada [17] (박하나, et al., 2021). the scope of review is limited to official curriculum for children and adolescent students. This paper aims to focus on programs explicitly targeting older adults, and make nearly comprehensive coverage of not only curriculums, but also all cases of real classes and events that took place.

2.3 Headline and Lead Based Event Detection in News Articles

Headline, lead and body section (main text) are three distinct elements of general online news articles [18] (Dai, Taneja, & Huang, 2018). Headlines and leads are often separated from the body section and are used as a key feature in various news analysis tasks [19, 20, 21] (León, 1997; Qian, Yu, et al., 2019; Gaozhao, Dongfang, 2021). The length of text in the body section is longer than that in the headline or lead, and thus the topics included in the body section is inherently noisy and heterogeneous [22] (Liu, Morstatter, Tang, & Zafarani, 2016). Meanwhile, the headline can be an abstract highlighting the main point of that article [23] (Nir, 1993). Leads emphasize grabbing the attention of the reader by summarizing the key event of the article [24] (Spark & Harris, 2011). Thus, instead of full text, headline and lead were used as the main feature of article to determine its relevancy to the event of interest in our research.

3 METHODS

3.1 Preliminary Search

News articles published online in Korean by Korean news media since 2020.01.01 and before 2024.11.01 were searched from Bigkinds database [25] (Bigkinds). Report of the digital literacy education event for older adults have three essential semantic components; "digital literacy ", "education", and "older adults". These were conceptualized as three classes of search keyword. Search news articles that contain at least one keyword from each three classes. For the preliminary search, only one keyword for each class was chosen manually by researcher as shown in Table 1. Keyword classes are extended further research steps.

Each new article is previously classified to exactly one news category by Bigkinds. Based on this, articles categorized as "politics" (783), "international" (235), and "sports" (11) were excluded as they are very likely to be irrelevant to the event of the research interest. Table 0 shows data collected from Bigkinds. Named entities extracted by Korean BERT based named entity recognition model KPF-BERT-NER [26] were collected along with keywords, article texts and metadata.

Table 1. Preliminary Search Keywords

Class	Keyword
Digital Literacy	디지털 (digital)
Education	교육 (education)

Class	Keyword
Older Adult	노인 (older adult)

Table 2. Data Items Collected From Bigkinds Database

Class	Items
Article	title, body section (first 200 characters)
Metadata	id, date, media, author, URL
Named Entity*	person, location, organization
Keyword**	keyword

*Extracted by KPF-BERT-NER [26], **All nouns except number, email address and time words

3.2 Relevant Article Identification

Articles are divided into sentences to go through search for relevant headlines and lead. Headline and lead sentences describing the event of digital literacy training for older adults were identified by searching sentences that contains at least one keyword from each three classes. This is the same strategy used in article-level search. An article was defined as relevant and included for further analysis only if it included at least one relevant headline or lead.

3.3 Search Keyword Identification Based on Sentence-BERT embedding

For each of three keyword classes, additional keywords were identified. Additional keywords should have same semantics and in-sentence context with the original keywords.

Dataset of sentences containing only a pair of three initial keywords; (디지털, 교육), (디지털, 노인), and (교육, 노인); were built from the Bigkinds database. Every sentence in the beginnings of the articles in these datset were encoded by Sentence-BERT [27] (Reimers, N. 2019) and embeddings were built for all tokens. Pre-trained Sentence-BERT captures both token's semantics and in-sentence context as embeddings in its last hidden layer. Three BERT models; paraphrase-multilingual-mpnet [27] (Reimers, N. 2019), KoBERT [28] and KPFBERT [29] were used to encode the sentences. paraphrase-multilingual-mpnet and KoBERT are previously known to have good comprehension of Korean sentences. KPFBERT is expected to capture the contexts in Korean newspapers well as it is pretrained with Korean newspaper corpus.

Top 20 tokens with highest similarity with each of the manually chosen keyword were listed with three models. For instance, in dataset containing sentences with keywords "디지털(digital)" and "교육(education)", tokens having highest similarity with keyword "노인(older adult)" were listed. This aims to identify with which words the journalist commonly substitutes the word "노인(older adult)". Tokens and their source sentences were screened by the researcher to exclude semantically irrelevant cases. Then, the process of article search and headline/lead search was repeated with the extended keyword classes.

3.4 Event Identification by Article Clustering

Group of articles reporting the same instance of digital literacy training event are identified by clustering. As the articles are searched by three specific semantic components, semantic similarity between two news articles in the dataset is relatively high even if they report different events. Events are distinguished by unique name, location, stakeholders and responsible organizations of the educational program rather than the semantics of reporting sentences. Based on this logic, set of named entities were used as a feature for clustering instead of BERT-based embedding of full sentences.

Articles were represented by set of named entities representing location, organization, and business. Each article is vectorized using count vectorizer and clustered using hierarchical agglomerative clustering algorithm. Named Entities are previously extracted and provided by Bigkinds Database using KPF-BERT-NER.

3.5 Organizational Named Entity Clustering

Of the named entities provided by Bigkinds, business names are manually separated from names of other organizations. Then, the set of preprocessed organizational named entities were clustered to identify which groups of organizations act in the organizational ecosystem around digital literacy training on older adults. Names were encoded using KPFBERT and HDBSCAN clustering algorithm was applied. Regional prefixes were removed from the organizational named entities in advance to avoid noise in clustering.

4 RESULTS

4.1 Preliminary Search

After excluding 297 Articles categorized as “politics” (269), “international” (28) or “sports” (0), 6,108 articles were retrieved by preliminary search in Bigkinds database. Duplicates identified by Bigkinds system were excluded and 5,937 articles were included in the final dataset.

4.2 Relevant Article Identification

Table 3 shows number of relevent lead sentences and corresponding articles. Titles and first 200 characters of article bodies were reconstructed into dataset of 13,725 sentences of 5,937 articles. Among them, 132 sentences were identified as lead sentences reporting the event of interest in 112 articles.

Table 3: Search Results of Lead Sentences and Articles

	Preliminary search	Main search (with Sentence-BERT based extended keyword set)
Keywords	디지털 (digital) AND 교육 (education) AND 노인 (older adult)	(디지털 (digital) OR 스마트폰 (smartphone) OR 정보화 (digitization) OR 키오스크 (kiosk)) AND (교육 (education) OR 학습 (learning) OR 교실 (classroom) OR 수업 (class) OR 배우 (study; morpheme) OR 가르쳐 (teach; morpheme) OR 프로그램 (program), 수강 (attending class) OR 특강 (special lecture)) AND (노인 (older adult) OR 어르신 (elderly) OR 노년 (old population) OR 고령 (old age) OR 노령 (old age) OR 실버 (silver) OR 시니어 (senior) OR 경로당 (senior center))
Articles	5,937	23,738
Sentences	13,725	54,213
Relevant headline/lead	132	1,680
Relevant Articles	112	1,249

4.3 Search Keyword Identification Based on Sentence-BERT embedding

3 keywords in class digital literacy, 8 keywords in class training, and 7 keywords in class older adult were identified based on Sentence-BERT embedding similarity with the initial keywords. All three models contributed to the keyword identification. Model paraphrase-multilingual-mpnet was less robust in tokenizing Korean sentences than the models specialized in this language. However, it still produced meaningful embedding and similarity data to identify proper substitutes of the initial keywords. Korean is a morpheme-based language and two of the identified keywords were in form of morpheme rather than a complete word (배우-; study. 가르쳐-; teach). These are identified with model paraphrase-multilingual-mpnet and KPFBERT.

Table 4: Top 20 Tokens from the Articles with highest Sentence Bert Embedding Similarities with Three Search Keywords

Model	디지털 (digital)	교육 (education)	노인 (older adult)
paraphrase-multilingual-mpnet	량 (역량), 역 (역량), 바 (이른바), 환 (일환), **ICT**, **화 (정보화; digitization)**, 온라인, 일 (일환), 활동, 법을, 들의, 내용, 세요, 스 (서포터스), 전환, 움 (배움), 과학기술, 대한, 필 (필수), **및** (**스마트폰** 및 **키오스크**; **smartphone** and **kiosk**)	**배 (배움; learning)**, **학습 (learning)**, 과정, 과정, 받는, 르, **교실 (classroom)**, **수업 (class)**, **훈련 (training)**, **학습**, **배우 (study; morpheme)**, 가는, **움 (배움)**, 해주는, **쳐 (가르쳐; teach; morpheme)**, 스쿨, 학 (장학), 이번, 춰 (발맞춰), 대상	**어르신 (elderly)**, **노 (노년; old population)**, **년 (노년층)**, **령 (고령층; old age)**, 세 (60 세), **고 (고령층)**, 들, 들을, 들의, 세대, 보는, 복지, 을, **퇴 (퇴직; retirement)**, 한 (스마트한), 들에게, **층 (고령층)**, 물 (범물), 마 (마사회), 관 (복지관),
KoBERT	온라인, 이른바, 다양한, 심리, 일자리, **스마트폰**, 역량, 된, 지원센터, 맞춤형, 사용, 고객, 핵심, 천안, 라이프, **정보 (정보화)**, 운영하는, 에도, 지원을, 주문	금융, 과정, **프로그램 (program)**, **훈련**, **수업**, 대상, 센터, 지원을, 주제로, 하는, 다양한, 홍보, 이동, 능력, 인기, 서비스, 변화, 사회공헌, 환경, 해주는	우리은행, 들, **층 (고령층)**, 음식, 교사, 복합, **고령**, 대상, 주민, 시민, 청년, 인천시, IT, 실, 이른바, 필요한, 신한은행, 일자리, 정책, 행복
KPFBERT	노령, 스트레스, 사용법, 세대, 극복, 경주시, 튜, ##포터, ##정보, 동해, ##터, **단말기 (terminal)**, 광진구, 함평, 군포시, ##케어, 노년층, 대화, ##에도, 예천군	**배우**, 강화, 역량, 활성, 과정, ##주, ##화, 협력, 사업, ##강, **수강 (attending class)**, **특강 (special lecture)**, 금융, ##행, 장소, **에듀**, **교실**, ##스크, **가르쳐**, 앞둔	**노령 (old age)**, ##세대, 노 (노노케어), **어르신**, 정보, 함평, ##케어, 예천군, 스마트폰, **고령화**, 스트레스, ##포터, **실버 (silver)**, 군포시, **시니어 (senior)**, 기기, **경로당 (senior center)**, 한전, 프로그램, 무료

*Keywords with relevant semantics are highlighted in **bold**. Keywords selected after validating the source sentence are **underlined**.

Table 5 validates the strategy of identifying relevant lead/headline using composition of three keyword sets. 92.31% of retrieved sentences were describing the event within research interest and has contributed to identifying relevant articles.

For the main search, accuracy of final task; article identification' was proposed instead of sentence identification due to the large number of sentences.

Table 5. Validation of Headline/Lead Identification and Identification of Stop Words

Item	Preliminary search	Additional search (based on word similarity)
Keyword Matches	143	1,680
Relevant Headline/Lead	132	-
Accuracy	92.31%	93.59% (*)
Stop Words(Identified from preliminary search)	복지 예산 (welfare budget), 삭감 (budget cut), 인권위 (Human Rights Commission (*abbreviation), 주장 (claim), 스마트경로당 (smart senior center (*project name), 슈퍼브레인 (SUPERBRAIN (*product name))) (6)	

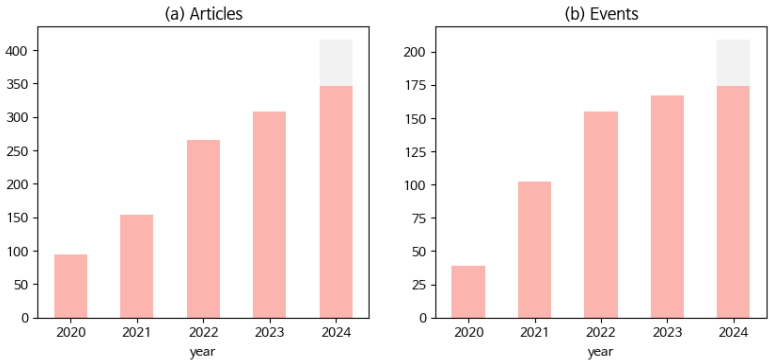
*Accuracy of relevant article identification. Lead/headline sentences were not validated due to the large size of dataset.

4.4 Event Identification by Article Clustering

528 clusters of similar articles were identified by agglomerative clustering. Clusters were manually refined to 603 distinct events of digital literacy trainings.

Figure 1. shows the annual trend of the collected articles and reported events. Number of events have rapidly, and steadily increased from 2020 to 2024. There's distinctive sharp in number of events increase from 2020 to 2021.

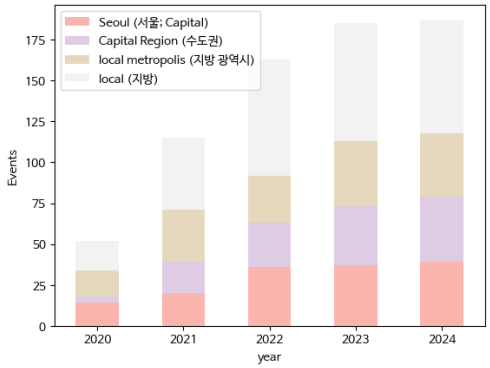
Figure 1. (a) Number of Articles Reporting digital literacy training sessions for older adults in South Korea. (b) Number of corresponding events. Gray area shows the estimated value for November and December of 2024.



4.5 Regional Distribution

Figure 2. shows the regional distribution of annually reported events. The overall increasing trend is evenly distributed among Seoul (Capital) and local area. More events are reported in local areas than in capital region. Especially, local rural areas contribute to the largest number of events since 2021.

Figure 2. Number of reported digital literacy training events for older adults in South Korea by regional bases

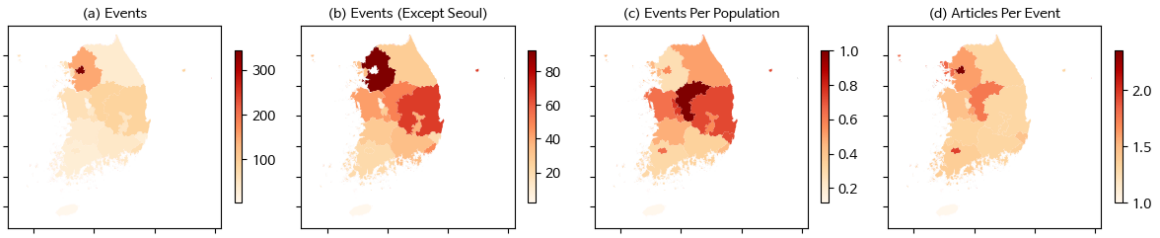


The regional distribution of reported digital literacy events are plotted on the map of South Korea in Figure 3. Most of the events are based on Seoul. Other than Seoul, Gyeonggi-do (경기도), Gyeongsangbuk-do (경상북도), Chungcheongbuk-do (충청북도) and Chungcheongnam-do (충청남도) are the provinces with largest number of events.

Regarding the population, Chungcheongbuk-do (충청북도), Gyeongsangbuk-do (경상북도) and Chungcheongnam-do (충청남도) has largest number of events per population.

Average number of articles reporting a single event was also analyzed to study the regional distribution of press release and journalism. Events based on Seoul are most repeatedly reported by media. Events based on Chungcheongbuk-do (충청북도) and Gyeonggi-do (경기도) are also often repeatedly reported.

Figure 3. (a) Number of reported digital literacy training events for older adults in South Korea by regional bases. (b) Number of events except Seoul (Capital). (c) Number of events per population. (d) Average Number of Articles reporting one event.

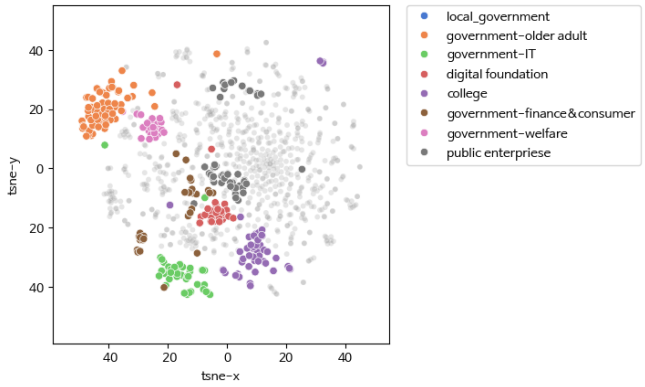


4.6 Organizational Ecosystem

Among the organizations that host and support the digital training sessions, local governments and businesses were manually identified and extracted. Besides these two, 17 more organizational classes were identified by HDBSCAN clustering algorithm; 1.government-older adult, 2.government-IT, 3.digital foundation, 4.college, 5.government-finance&consumer, 6.government-welfare, 7.public enterprise, 8.local community, 9.lifelong education, 10.education, 11.library, 12.Samsung (business group), 13.healthcare, 14.volunteer work, 15.veterans affairs, 16.welfare for the disabled and 17.senior center.

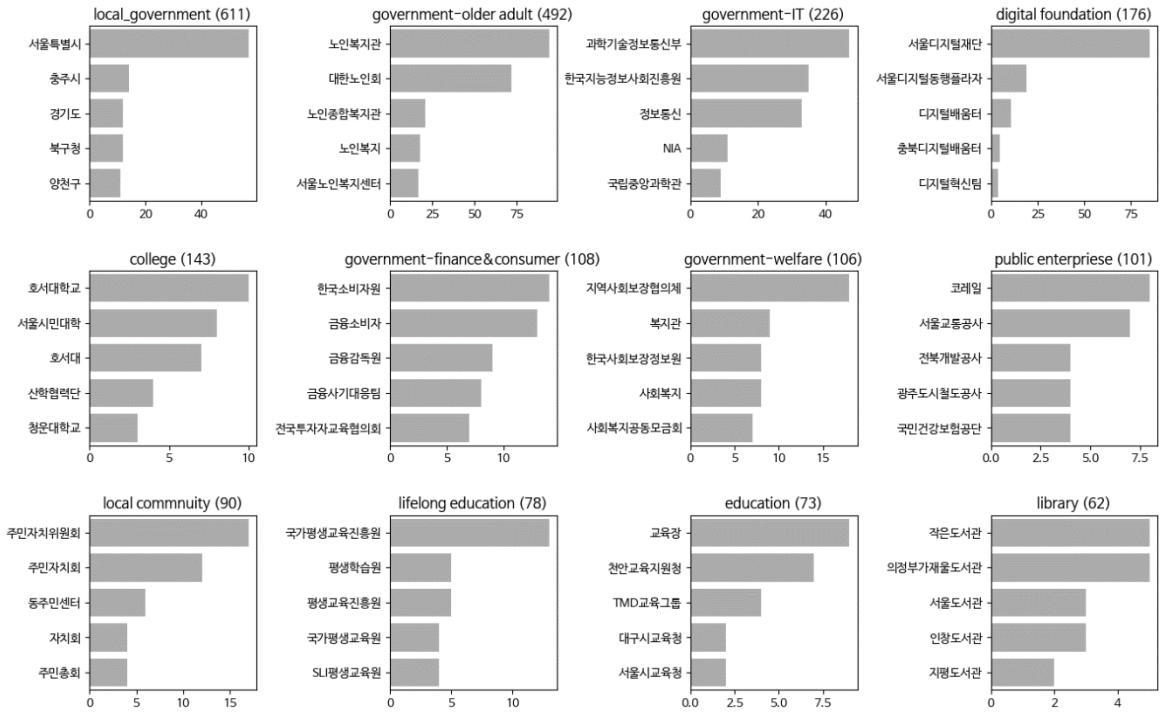
Clusters corresponding to the classes are displayed in Figure 4. Each data point represents one named entity describing organization, and its position is a reduced representation of the named entity's BERT-based embedding

Figure 4. Reduced representation of organization named entities' BERT embedding and identified clusters. Only 8 largest clusters are displayed. Businesses (corporations) are not included.



12 largest classes of the organizations and the most frequent named entities are listed in Figure 5. Local governments and government agencies working for older adults consist the largest classes. Following are government agencies and foundations working on IT and digitization, college, and other government agencies.

Figure 5. 5 Most frequent examples of organizational named entities in the organization clusters. In the parenthesis are the total numbers of appearances of the named entities in each cluster. Businesses (corporations) are not included.



5 CONCLUSION

The main contribution of this paper is the news-based database of recent digital training sessions conducted in South Korea for older adults. The database is retrieved with keywords identified with Sentence-BERT encoding. This approach may contribute to minimizing unsearched data and avoid selection biases rooted in human selection of search keywords. The final dataset provides information about 603 distinct events reported in 1,249 articles. The database provides opportunity for review in various aspects, which could possibly conclude in meaningful feedbacks for planning of new digital literacy programs and policy making.

The number of training events have been increasing every year since 2020 to 2024. It has sharply increased from 2020 to 2021, and this trend could relate to the quarantines due to the pandemic of COVID-19 started at March of 2020. Relationship could be proved in future work. While large proportion of the events are based on Seoul, the reported events are evenly distributed in Seoul and local areas regarding the populations. Press release and media deliveries tend to focus on the events held in Seoul and capital regions. Chungcheongbuk-do (충청북도) shows distinctively large number of reported events. Evaluation of local government policies in this province may give some insights for agencies in other local areas.

Organizational ecosystem in South Korea around digital literacy training for older adults is identified from the database. Different sectors of government agencies are contributing for the progress, and other actors including business, college, and local communities are also identified. Trainings in South Korea is mainly governed by local governments. Government sectors working for older adults and IT promotion are also important groups.

This paper focuses on building a large qualified database and conducting macroscopic analysis of the trend and ecosystem of digital literacy trainings for older adults. Our future work plans to extract and analyze details; location, duration, frequency, size of class and tutors of each training sessions. Textbook and other educational resources would be archived and attached to the database. In long term, researches could study our database to identify meaningful patterns that should be applied, or be avoided in the next generation of digital literacy programs for older adults.

REFERENCES

[1] OECD (2001), "Understanding the Digital Divide", OECD Digital Economy Papers, No. 49, OECD Publishing, Paris, https://doi.org/10.1787/236405667766.

[2] Rogers, S. E. (2016). Bridging the 21st century digital divide. TechTrends, 60(3), 197-199.

[3] Lythreatis, S., Singh, S. K., & El-Kassar, A. N. (2022). The digital divide: A review and future research agenda. Technological Forecasting and Social Change, 175, 121359.

[4] Dewan, S., & Riggins, F. J. (2005). The digital divide: Current and future research directions. *Journal of the Association for information systems*, *6*(12), 298-337.

[5] Hoffman, D. L., Novak, T. P., & Schlosser, A. (2000). The evolution of the digital divide: How gaps in Internet access may impact electronic commerce. *Journal of computer-mediated communication*, *5*(3), JCMC534.

[6] Hargittai, E. (2000). Second-level digital divide: differences in people's online. *Retrieved from Skillschnm. gmu. edu/digitalhistory/links/pdf/introduction/0.26 c. pdf*.

[7] Van Dijk, J. A. (2005). M. The deepening divide: Inequality in the information society. *Mass Communication & Society*, *11*(2), 221-234.

[8] Scheerder, A., Van Deursen, A., & Van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and informatics*, *34*(8), 1607-1624.

[9] Mossberger, K. (2003). *Virtual inequality: Beyond the digital divide*. Georgetown University Press.

[10] Van Deursen, A., & Van Dijk, J. (2011). Internet skills and the digital divide. *New media & society*, *13*(6), 893-911.

[11] Blažič, B. J., & Blažič, A. J. (2020). Overcoming the digital divide with a modern approach to learning digital skills for the elderly adults. *Education and Information Technologies*, *25*, 259-279.

[12] Martin, A. (2008). Digital literacy and the "digital society". *Digital literacies: Concepts, policies and practices*, *30*(151), 1029-1055.

[13] Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in human behavior*, *72*, 577-588.

[14] United Nations. (2012). *E-Government Survey 2012. United Nations E-Government Survey 2012*. New York: United Nations.

[15] 김학실, & 심준섭. (2020). 노인의 디지털 리터러시와 사회활동. *정책분석평가학회보*, *30*(2), 153-180.

[16] 과학기술정보통신부, 한국정보화진흥원. (2018). 2018 디지털정보격차 실태조사 보고서. South Korea.

[17] 박하나, 진명화, 박지우, & 임규연. (2021). 국내외 디지털 리터러시 교육과정 분석. *교육문화연구*, *27*(5), 75-101.

[18] Dai, Z., Taneja, H., & Huang, R. (2018, August). Fine-grained structure-based news genre categorization. In *Proceedings of the Workshop Events and Stories in the News 2018* (pp. 61-67).

[19] León, J. A. (1997). The effects of headlines and summaries on news comprehension and recall. *Reading and Writing*, *9*, 85-106.

[20] Qian, Y., Deng, X., Ye, Q., Ma, B., & Yuan, H. (2019). On detecting business event from the headlines and leads of massive online news articles. *Information Processing & Management*, *56*(6), 102086.

[21] Gaozhao, D. (2021). Flagging fake news on social media: An experimental study of media consumers' identification of fake news. *Government Information Quarterly*, *38*(3), 101591.

[22] Liu, H., Morstatter, F., Tang, J., & Zafarani, R. (2016). The good, the bad, and the ugly: uncovering novel research opportunities in social media mining. *International Journal of Data Science and Analytics*, *1*, 137-143.

[23] Nir, R. (1993). Discourse analysis of news headlines. *Balšanwt iybriyt*, (37), 23-31.

[24] Spark, D. B., & Harris, G. (2010). Practical newspaper reporting.

[25] 한국언론진흥재단 (Korea Press Foundation). 빅카인즈 (BigKinds). www.bigkinds.or.kr

[26] 빅카인즈랩. (2022). BIGKINDS-LAB. GitHub Repository. Available: https://github.com/KPF-bigkinds/BIGKINDS-LAB

[27] Reimers, N. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv preprint arXiv:1908.10084*. Available: https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2

[28] SK T-Brain. (2019). KoBERT. GitHub Repository. Available: https://github.com/SKTBrain/KoBERT

[29] 한국언론진흥재단. (2021). kpfbert. GitHub Repository. Available: https://github.com/KPFBERT/kpfbert