

Theme-Related Keyword Extraction from Free Text Descriptions of Image Contents for Tagging

Joonmyun Cho*, Yoon-Seop Chang*, Sung-Ho Lee*

*Hyper-connected Communication Research Laboratory, ETRI(Electronics and Telecommunications Research Institute),
218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700, Korea.

jmcho@etri.re.kr, ychang76@etri.re.kr, sholee@etri.re.kr

Abstract— This paper discusses a method for automatic theme-related keyword extraction from users' natural language comments on their photographs and videos. 'Theme' indicates the concepts circumscribing and describing the content of the photos and videos such as pets, natural sites, palaces and places. The method employs a deep learning algorithm, RNN(Recurrent Neural Network) that is good at recognizing implicit patterns of sequential data. The method has been applied to the construction of a place-related image content DB, and delivers reasonably good performance even in case the measure (i.e. themes of image contents) is abstract and vague.

Keywords— Image Content DB, Keyword Tagging, Content Search, Tag Extraction, Recurrent Neural Network (RNN)

I. INTRODUCTION

Let's imagine an automatic construction of theme-related image content DB using photographs and videos collected from various social media such as SNS and blogs. Here, 'theme' indicates concepts circumscribing and describing the content of the photographs and videos such as pets, palaces, natural sites and places. Such theme-related image content DB can be utilized in various applications. For example, imagine the task of choosing places to take scenes of a movie or TV drama. The person in charge of the task in a film-making project can refer to the photos or videos retrieved from a place-related image content DB in order to reduce the number of candidate places which he/she should go out to and check the real conditions or feelings of. By checking the real conditions only of a small number of candidate places rather than many possible places, he/she can save a lot of time and money. We can imagine many other tasks or applications that could benefit from such theme-related image content DBs such as choosing travel destination, determining a pet dog breed and so on.

Those image content DBs should provide an easy and quick way of finding the image contents that are related closely to users' task. The most familiar way of searching is to use search keywords. In order for the DB systems to provide keyword search, each photos and videos have to be tagged with keywords when they are stored in the DBs. This paper discusses an automatic theme-related keyword extraction from

users' natural language comments on their photos and videos; in most social media, user comment texts are posted together with the image contents. The theme-related keyword extraction system employs a deep learning algorithm, RNN(Recurrent Neural Network) that is good at recognizing implicit patterns of sequential data like sequence of words in user comments. In existing methods having pre-set candidate keywords for a certain target domain (i.e. theme), many legitimate keywords as closely associated with the theme can be omitted or many unnecessary keywords as weakly associated with the theme can be extracted, so that the quality of extracted keywords tends to be worsened. This is because the concept of the theme of image contents is vague so that it cannot be defined with explicit rules and that even the same keyword can have different meanings or nuance in the context of a user comment.

II. BACKGROUND TECHNOLOGY AND RELATED WORK

A. Social Media

Huge amounts of contents are posted and distributed through various social media such as blogs and SNS(Social Network Service). Especially, owing to increase of the Internet bandwidth and progress of telecommunication technology, the amounts of image contents such as photographs and videos are growing explosively. Most social media platforms provide open APIs for users and computer systems to use the image contents freely and easily [1], [2]. Looking at typical social media such as flickr[3] and facebook[4], we can see that photos and videos are posted together with user comments written in natural language. Thus, if we analyse the user comments we can find the sort (i.e. theme) of the image contents and, furthermore, we can extract proper keywords related to the image content. Figure 1 shows a snap shot of a flickr web page posting a photo and user comments on it.

There are numbers of work such as Rae[2] and Kim[5] that extract certain information from unstructured data like natural language text. Kim developed a system that extracts from email text the information related to a meeting such as the name, place, date and time of the meeting. The system

employs CRF(Conditional Random Field) machine learning algorithms to recognize such information words in the texts. Rae et al. also utilize CRF algorithms to extract from user comment text the information related to a place where the flickr image content are taken.

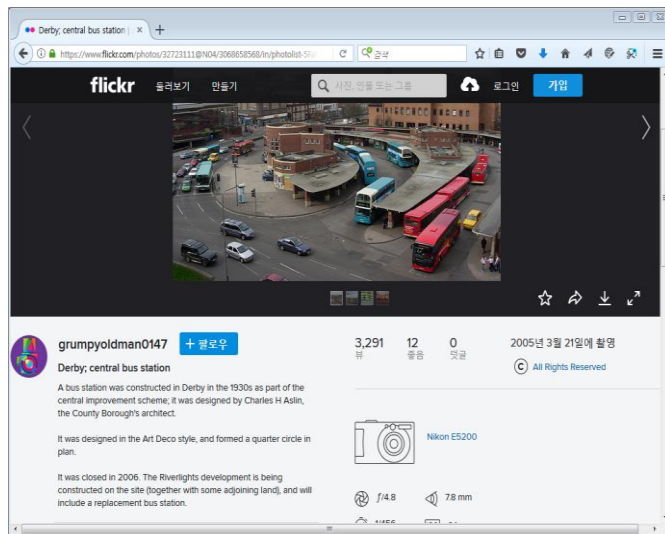


Figure 1. A snap shot of a flickr web page posting a photo and user comments on it

B. Recognition of Patterns of Theme-Related Keywords

Extraction of keywords relating to a certain concept (i.e. theme) from user comments of natural language text corresponds to recognition of implicit patterns of unstructured data. The patterns of theme-related keywords are implicit because the boundaries of a theme cannot be precisely defined and thus, there are no explicit rules to separate proper keywords from improper ones. Actually, even people differ from one another in determining right keywords depicting features of, for example, places, pets and buildings in images. Nonetheless, the patterns certainly exist because humans, on the other hand, feel no difficulty performing the task and generally arrive at a number of keywords agreeable to everyone.

Various machine learning techniques are now being applied to the implicit pattern recognition. However, in case of natural language text, the patterns of a keyword do not involve only the attributes of the keyword itself. The order and/or sequence of words in the text also constitute the patterns. In other words, other words around the keyword and the relations between them compose a large portion of the patterns. Machine learning algorithms that well recognize such sequential patterns include HMM(Hidden Markov Model) and CRF [5], [6]. Recently, it is reported that RNN(Recurrent Neural Network), one of the deep learning algorithms, delivers better performance on sequential pattern recognition than HMM and CRF algorithms [7].

C. Machine Learning

In order to apply machine learning algorithms to the pattern recognition over data, we have to feed attributes of the data

and desired patterns to be recognized [6], [7]. As the attributes of words in the user comment text, besides the order of the words, we can consider various language features such as POS (Part of Speech), dependency relations and semantic categories [1], [2], [5]. Words assigned to the same POS generally display similar behaviour in terms of syntax, and words in a sentence connect with each other directly or indirectly. Moreover, a word represents some things or concepts so that it can be classified into categories of meanings according to its representations. Humans, in fact, unconsciously refer to such features when understanding sentences.

Meanwhile, machines do not know which patterns are significant and thus which patterns they should recognize. So, we have to inform the machine of the desired patterns by providing labels. While the language features as the attributes of data are usually provided and tagged automatically by natural language processing toolkits, the labels are provided and tagged manually by human users.

III. THE THEME-RELATED KEYWORD EXTRACTION

The theme-related keyword extraction consists mainly of three systems: *theme-related keyword pattern learning system*; *theme-related keyword extraction system*; and *language feature tagging system*. The theme-related keyword pattern learning system generates a model for pattern recognition of the keywords relating to a specific theme. The system is trained using machine learning data composed of word order, POS, dependency relations, semantic categories and labels. The theme-related keyword extraction system extracts keywords from any user comments based on the recognition model learned. The language feature tagging system automatically tags language features to each word in the user comments. The tagged user comment data is used as the training data off-line which is the input to the learning system and as pattern recognition data on-line which is the input to the extraction system. Figure 2 illustrates the process of the theme-related keyword extraction and location of each system in the process.

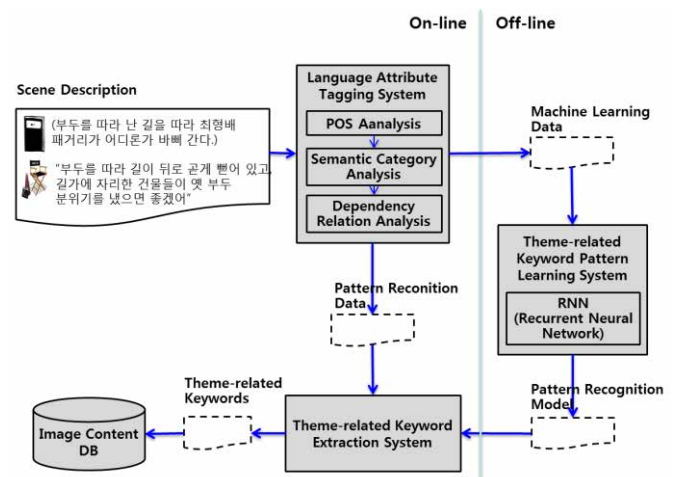


Figure 2. The theme-related keyword extraction process and its constituents

A. Tagging the Language Features

The language feature tagging system decomposes an input user comment text into syntactic units (i.e. words) preserving the original order of the words. And then, it tags language features such as POS, dependencies and semantic categories to each word. The system has been implemented using UWordMap[8] and UTagger[9] developed by Ulsan University, Korea. UWordMap is a Korean word map constructed based on the Standard Korean Dictionary of National Institute of Korean Language. UTagger is a toolkit that tags POS, dependency relations and semantic categories of words using UWordMap and various natural language processing algorithms.

Source	History	
1	온통	MAG 10 O_SEQ
2	조롱빛	NNG 1 현상__04/NNG 빛/NNG O_SEQ
3	움결	NNG 1 작용__01/NNG 운동__02/NNG 모양__02/NNG O_SEQ
4	의	JKG 1 O_SEQ
5	광경	NNG 1 언어단위/NNG 어근__01/NNG O_SEQ
6	하	XSA 1 O_SEQ
7	...	ETM 1 O_SEQ
8	잔디	NNG 2 풀__02/NNG 여러해살이풀/NNG B_SEQ
9	와	JC 2 O_SEQ
10	이름	NNG 2 기호__01/NNG 명칭__02/NNG 인명__01/NNG 이름/NNG 명성__01/NNG
11	모르	VV 2 O_SEQ
12	...	ETM 2 O_SEQ
13	충	XPN 1 접사__04/NNG 접두사/NNG O_SEQ
14	전연	NNG 1 모양__02/NNG 상태__01/NNG O_SEQ
15	...	JKG 1 O_SEQ
16	꽃	NNG 3 물질__02/NNG 조직/NNG 생물__01/NNG 식물__02/NNG 사람/NNG 여자
17	...	XSN 3 접사__04/NNG 접미사/NNG O_SEQ
18	...	SF 3 O_SEQ
19	그	MM 1 O_SEQ
20	주변	NNG 1 끝__01/NNG 가장자리/NNG O_SEQ
21	...	JKO 1 O_SEQ
22	돌려싸	VV 2 O_SEQ
23	...	ETM 2 O_SEQ
24	태고	NNG 1 시기__04/NNG 옛날/NNG O_SEQ
25	적	XSN 1 접사__04/NNG 접미사/NNG O_SEQ
26	부터	JX 1 O_SEQ
27	의	JKG 1 O_SEQ
28	아름드리	NNG 1 모양__02/NNG 겉모양/NNG B_SEQ
29	나무	NNG 2 식물__02/NNG 여러해살이식물/NNG 목재/NNG 재목__01/NNG 감__
30	와	JC 2 O_SEQ
31	그	MM 1 O_SEQ
32	한쪽	NNG 1 곳__01/NNG 방향__01/NNG O_SEQ
33	으로	JKB 1 O_SEQ
34	보이	VV 2 O_SEQ
35	...	ETM 2 O_SEQ
36	작	VA 1 O_SEQ
37	...	ETM 1 O_SEQ
38	언뜻	NNG 3 물__01/NNG 못__03/NNG 곳__01/NNG O_SEQ
39	까지	JX 3 O_SEQ
40	...	SP 3 O_SEQ
41	그	MM 1 O_SEQ
42	위	NNG 1 자리__01/NNG 위치__01/NNG 공간__05/NNG 곳__01/NNG 방향__01
43	로	JKB 1 O_SEQ
44	나비	NNG 3 하등동물/NNG 벌레__01/NNG O_SEQ
45	...	XSN 3 접사__04/NNG 접미사/NNG O_SEQ
46	이	VCP 3 O_SEQ
47	머	EC 3 O_SEQ
48	장자리	NNG 2 세간__01/NNG 살림살이/NNG 공간__05/NNG 자리__01/NNG 방사__04
49	...	XSN 2 접사__04/NNG 접미사/NNG O_SEQ
50	이	JKS 2 O_SEQ
51	...	NNG 1 말__01/NNG 단어/NNG 연계__03/NNG 사이__01/NNG O_SEQ
52	...	JKO 1 O_SEQ
53	것	VV 1 O_SEQ
54	어	EC 1 O_SEQ
55	...	VV 0 O_SEQ
56	...	EC 0 O_SEQ
57	...	SE 0 O_SEQ

Figure 3. An example of machine learning data composed of language feature tags and labels

Figure 3 shows an example of machine learning data the theme-related keyword pattern recognition. The data preserves the order of words and each word is tagged with language features and labels. The tags such as NNG, JKG and XSA denote POS features of the word, meaning a common noun, a

genitive case marker and an adjective-derived suffix, respectively. The tags such as 1, 2 and 3 denote word dependencies, meaning the distance to other word the word depends on. The semantic categories are marked with such tags as 작용__01, 운동__02 and 여자__02, meaning action, movement and female, respectively. Such language feature tags are given automatically by UWordMap and UTagger.

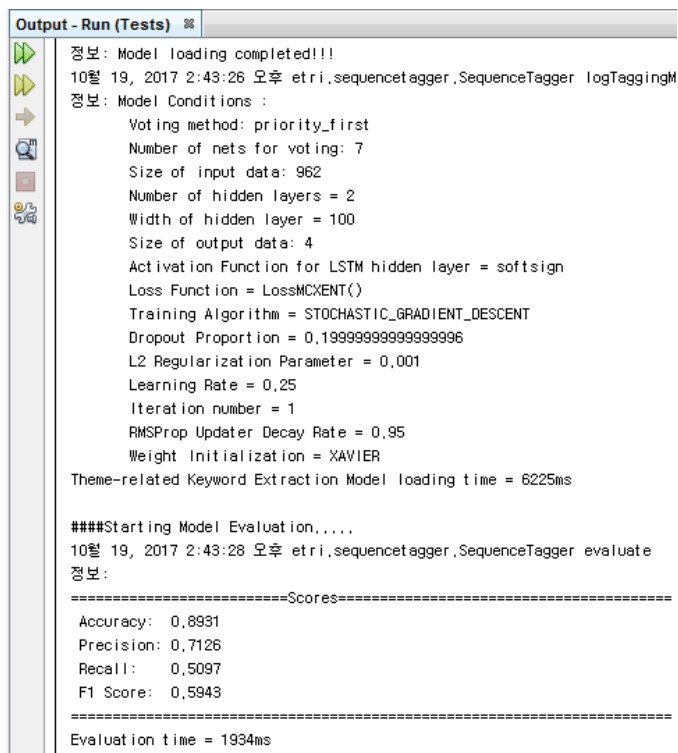
Note that those language attributes tagged might include many noises i.e. wrong tags. In other words, because the natural language processing algorithms themselves are also based on statistics model and/or machine learning techniques, their analysis might be inaccurate. Moreover, a word can have many different meanings (i.e. semantic categories) in different texts (i.e. sentences) according to the context of the sentences. The system, however, does not tag the right semantic categories of the word in a sentence but tags all possible semantic categories of the word. In other words, the semantic category in this paper does not mean the exact meaning of the word in a user comment text. It is because judging accurate meanings of a word in a sentence is very difficult for the state-of-the-art techniques or at least, the accuracy is so poor that much noise has to be included inevitably.

The last tags of each word are the labels which are given by human users as the answer of the pattern recognition. The labels are tagged according to IOB2 tagging model [10]. In this model, B tag denotes the beginning of a target pattern (i.e. a theme-related keyword), I tag denotes the inside of the target pattern, and O tag denotes the outside of the target pattern. So, with these label tags, we can separate the target theme-related keywords from others.

B. Learning Patterns of the Theme-Related Keywords

The theme-related keyword pattern learning system is trained to generate a pattern recognition model using a bulk of data (i.e. machine learning data) which incorporates the language features tags and labels. This paper employs RNN algorithms in order that the system can learn a model for sequential data patterns. As a toolkit for RNN algorithms, DeepLearning4J[11] is used. DL4J is developed using Java so that it is easy to integrate DL4J with application system.

The RNN used in this paper consists of 4 neural network layers. The first layer is the input layer and has 962 sigmoid neurons (i.e. nodes) according to the number of input attributes of data. The second and third layers are the hidden layers having 100 LSTM(Long and Short-Term Memory) neurons respectively. Lastly, the fourth layer is the out layer of 4 softmax neurons for the three label tags and an exceptional output tag; the exceptional case is thought to occur by user's false label tagging. SGD(Stochastic Gradient Descent) algorithm is used for the training of the neural network. And, as the cost function of the training algorithm, LossMCXENT, a sort of the negative log-likelihood cost function is used. This paper applies the Dropout regularization and L2 regularization. Such RNN explained above can be thought quite typical one having no special features. In Figure 4 we can see the network configuration and some hyper-parameters of the network.



```

Output - Run (Tests)
정보: Model loading completed!!!
10월 19, 2017 2:43:26 오후 etri.sequencetagger.SequenceTagger logTaggingM
정보: Model Conditions :
  Voting method: priority_first
  Number of nets for voting: 7
  Size of input data: 962
  Number of hidden layers = 2
  Width of hidden layer = 100
  Size of output data: 4
  Activation Function for LSTM hidden layer = softsign
  Loss Function = LossMCKENT()
  Training Algorithm = STOCHASTIC_GRADIENT_DESCENT
  Dropout Proportion = 0.19999999999999996
  L2 Regularization Parameter = 0.001
  Learning Rate = 0.25
  Iteration number = 1
  RMSProp Updater Decay Rate = 0.95
  Weight Initialization = XAVIER
Theme-related Keyword Extraction Model loading time = 6225ms

####Starting Model Evaluation....
10월 19, 2017 2:43:28 오후 etri.sequencetagger.SequenceTagger evaluate
정보:
=====Scores=====
Accuracy: 0.8931
Precision: 0.7126
Recall: 0.5097
F1 Score: 0.5943
=====
Evaluation time = 1934ms

```

Figure 4. The RNN configuration with some hyper-parameters and evaluation results based on test data

C. Extracting Theme-Related Keywords

The theme-related keyword extraction system performs its task based on the pattern recognition model learned. In fact, the system applies in parallel 7 models trained with the same training data. And, the pattern recognition results (i.e. resulting label tags) from the 7 models are synthesized into final one. Specifically, the priorities of the label tags are set as B tag > I tag > O tag, and then the tag of the highest priority is finally decided as the tag of the word even if the tag has the least incidence among the 7 tags returned. This priority-based voting scheme expands, by taking account of any probability of theme-related keywords, the potential keywords that will be extracted. The reason for adopting such scheme is that the main use of the extracted keywords is search keywords so that they are encouraged to increase the recall rate of the search rather than the precision.

IV. EXPERIMENT AND EVALUATION

The method discussed in Section 3 has been implemented in a construction of place-related image content DB. The primary purpose of the DB is to support the task of choosing places to take scenes of movies or TV dramas. In order to prepare the machine learning data (i.e. the training data) for the place-related keyword pattern learning system, 663 sentences or passages are collected from various TV drama scripts that describe the conditions or features of the filming places. Among them, for example, is the passage: "It is Hyunwook's house secluded in quiet countryside paths. There is a humble but modern space of clear view with no walls, no gates. In a large garden are carelessly grown grass and wild

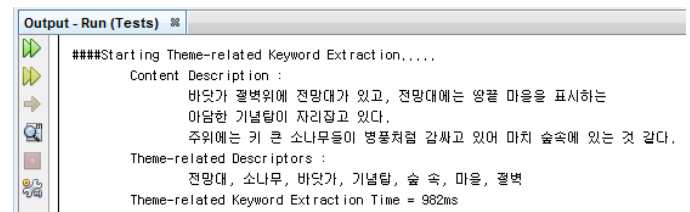
flowers...." The passages collected were tagged by the language feature tagging system into preliminary training data, and then human users marked the BIO labels on the preliminary training data into the final training data.

A. Place-Related Keyword Pattern Recognition Model

Using the machine learning data prepared, 7 RNNs were trained. 70% of the data corresponding to 464 passages out of the 663 passages was used as the training data and the rest of the data was used as the test data. Finally, 7 place-related keyword pattern recognition models were generated with the same training data, and they deliver average accuracy of 89% over the test data for the labels. In Figure 4, we can see the evaluation results of the models such as accuracy, precision and F1 score.

B. Place-Related Keywords for Image Contents

Various user comments on photos and videos are gathered from various media and processed by the place-related keyword extraction system equipped with the 7 pattern recognition models. Figure 5 shows one example of the processing. The user comment in the example is: "There is an observation platform on top of seaside cliff and in the platform is a small monument saying the village is the end of the earth. As tall pine trees surround the platform like a folding screen, feeling of being in forests occurs." The extracted keywords are: observation platform; pine tree; seaside; monument; forest; village; cliff. After processing over 130 user comments, the average performance has been estimated by human users at: precision: about 76%; recall: about 94%. In the example of Figure 5, the precision is 5/7 = 71.43% and the recall is 5/5 = 100% because human users judged the following 5 keywords are the right place-related keywords and they are all included in the extracted keywords: observation platform; pine tree; seaside; monument; cliff.



```

Output - Run (Tests)
####Starting Theme-related Keyword Extraction....
Content Description :
바닷가 절벽위에 전망대가 있고, 전망대에는 쌍굴 마을을 표시하는
아름다운 기념탑이 자리잡고 있다.
주위에는 키 큰 소나무들이 병풍처럼 감싸고 있어 마치 숲속에 있는 것 같다.
Theme-related Descriptors :
전망대, 소나무, 바닷가, 기념탑, 숲 속, 마을, 절벽
Theme-related Keyword Extraction Time = 982ms

```

Figure 5. An example of the place-related keyword extraction

V. CONCLUSIONS

This paper discusses a method to extract theme-related keywords automatically from user comments on photos and videos that describe the content of the image contents. The method employs RNN of 4 layers with 100 nodes per hidden layer to generate theme-related keyword pattern recognition model. One of the language attributes to compose the training data is semantic categories of words. The method tags all possible semantic categories of a word not exact one of the word in a user comment text. With the training data, 7 models are trained and they voted for the final label. The label of higher priority is selected among the 7 label tags returned by the 7 models regardless of its incidence. This priority-based

voting scheme increases the recall rate of search by expanding potential keywords that would be extracted. The method has been implemented in the construction of place-related image content DB. The place-related keywords extraction system renders an average precision of about 76% and an average recall of about 94% according to human user estimation for the keywords extracted.

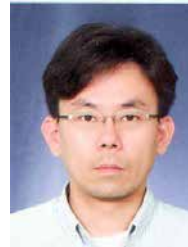
This paper shows it is possible to deliver reasonable performance in the automatic extracting of theme-related keywords when using machine learning technique, especially, RNN, even when the measure (in this paper, themes of image contents) is abstract and vague. Moreover, even in case the attributes used for the training data has some noise (in this paper, inaccurate semantic categories of a word with respect to the word's exact meaning in text), the accuracy of the extracted keywords is reliable. The method in this paper can also be used to determine the sort of photos and videos for classification by analysing user comments on the contents instead of analysing image contents themselves.

ACKNOWLEDGMENT

This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2017.

REFERENCES

- [1] S. Kumar, F. Morstatter, and H. Liu, "Twitter Data Analytics," *Database Management & Information Retrieval*, 2013.
- [2] A. Rae, A. Popescu, V. Murdock, and H. Bouchard, "Mining the Web for Points of Interest," in *Proc. Of International SIGIR Conference on Research and Development in Information Retrieval*, 2012.
- [3] (2017) flickr [Online]. Available: <https://www.flickr.com/>
- [4] (2017) facebook [Online]. Available: <https://www.facebook.com/>
- [5] K. R. Kim, "Location Extraction from Meeting Announcements," KAIST, Master's Thesis, 2012.
- [6] B. T. Jang, "Next-Generation Machine Learning Technologies," *Communications of the Korean Institute of Information Scientists and Engineers*, 2007.
- [7] A. Graves, "Supervised Sequence Labelling with Recurrent Neural Networks," *Studies in Computational Intelligence*, Springer, 2012.
- [8] Y. Bae, and C. Ock, "Introduction to the Korean Word Map(UWordMap) and API," in *Proc. of the 26th Annual Conference on Human and Cognitive Language Technology*, 2014.
- [9] J. Shin, and C. Ock, "Optional features for speeding up UTagger," in *Proc. Of the 24th Annual Conference on Human and Cognitive Language Technology*, 2012.
- [10] T. Ek, C. Kirkegaard, H. Jonsson, and P. Nugues, "Named Entity Recognition for Short Text Messages," in *Proc. Of International Conference of the Pacific Association for Computational Linguistics*, 2011.
- [11] (2017) Deeplearning4J [Online]. Available: <https://deeplearning4j.org/>



Joonmyun Cho received his B.S., M.S. and Ph.D. degrees in mechanical engineering from KAIST(Korea Advanced Institute of Science and Technology), South Korea, in 1993, 1995 and 2006, respectively. He joined ETRI(Electronics and Telecommunications Research Institute), South Korea in 2007 and was involved with the URC (Ubiquitous Robotic Companion) project until 2011 and Beyond Smart TV project until 2015. Dr. Cho is currently working in Intelligent IoT SW Platform project as a senior researcher. His research interests

include knowledge based systems, intelligent agent systems and machine learning.



Yoon-Seop Chang received his B.S., M.S., and Ph.D. degrees in geographic information system from Seoul National University, South Korea, in 1999, 2001 and 2005, respectively. He joined ETRI (Electronics and Telecommunications Research Institute), South Korea, in 2005 and is currently working as a principal researcher. Since 2008, Dr. Chang has also been a faculty member of University of Science and Technology, South Korea, as an associate professor. His research interests include geographic information

system, web mashup, augmented reality and virtual reality.



Seong-Ho Lee received his B.S. and M.S. degrees in computer science from Chungbuk National University, South Korea, in 1997 and 2000, respectively. Since 2000, he has been a senior member of research staff with ETRI, South Korea, and he is also working toward the Ph.D. degree in computer science Chungbuk National University. Mr. Lee is currently working in Location-based Smart Content Platform project as a senior researcher. His research interests are spatio-temporal database systems, geographic information systems, and location-based services.