International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.5 (2014), pp.317-326
http://dx.doi.org/10.14257/ijmue.2014.9.5.32

Location Recommendation System Using Big Data

Ki-Young Lee[1], Jeong-Jin Kang[2], Hye-Kyoung Ahn[1], Kyu-Ho Kim[1,*],
Gyoo-Seok Choi[3], Sung-Jai Choi[4] and Sun-Jin Oh[5]

[1]Department of Medical IT and Marketing, Eulji University, Seongnam, Korea
(kylee@eulji.ac.kr, sweety3003@hanmail.net, khkim@eulji.ac.kr)
[2]Department of Information and Comminication, Dong Seoul University, Seongnam, Korea
jjkang@du.ac.kr
[3]Department of Computer Science, ChungWoon University, Incheon, Korea
lionel@chungwoon.ac.kr
[4]Department of Electronic Engineering, Gachon University, Seongnam, Korea
csj0717@gachon.ac.kr
[5]Department of Computer and Information Science, Semyung University, Jecheon, Korea
sjoh@semyung.ac.kr
**Corresponding Author: khkim@eulji.ac.kr*

Abstract

Active SNS and the amount of data which increases very fast these days, at WEF, the first remarkable technology of arising 10th technology of 2012 is called Big Data. Big Data extracts the value of collected, saved, analyzed, and processed large data by using database management tool. In this paper, the data structured and unstructured on SNS has set its position through abstracting significant information from the Big Data technology. From this Big Data, more information can be achieved to set position by abstracting materials from unstructured data.

Keywords: *Big Data, Text Mining, SNS, Position*

1. Introduction

Today, we are living in the age of over zetabytes of information and an explosive increase of data due to increasing mobile phone or social network system, and etc. In this enormous data, if the existing data analysis used only structured data, the big data can handle and process unstructured data [1]. According to the MIT survey as a representative form of unstructured data, the mobile and social network service etc, there is an outbreak of atypical data which exceed 2.5*10^{18} byte and over 30 billion Facebook messages and 1 million Twits are made. The goal of Big Data which can analyze unstructured data is extracting enormous value by using this technique.

Moreover, it can create many things so it is directly connected to reinforce the age of national competitiveness. Increasing interests of Big Data stand on massive data, natural language processing, and machine learning techniques, such as automatic translation of Google and 'Siri' of Apple, is increased to understanding language flow and context reasoning services. And applying to reasoning services other new industry

ISSN: 1975-0080 IJMUE

fields is continuously developed [2-4]. The state also instituted the big data, with the goal of smart government realization, to apply it on info-communication, education, medical service, finance, and *etc*. Also helps decrease cost and makes a convenient government[5-6].

Analysis of commercial rights referred by structured data rather than unstructured data is used when selecting existing direction[7-8].

The system suggested in this paper, is a system to make information applied to analyze and process the vast data in the SNS. From this, users can extract important information of positioning not only by structured data but also unstructured and semi-structured data.

2. Related Research

2.1. Big Data

Big Data has a massive data size and includes diverse data of unstructured data such as photo or video; representative features are Volume, Various, and Velocity [9]. Big Data includeS not only existing structured data but also enormous unrefined unstructured data or semi-structured data. Before, data was mainly used as accumulation and sharing, now data is mainly used as data analysis and prediction. In this, unstructured data refers to available text analysis of text, image, and video data which is not saved in the fixed field. Semi-structured data is saved in the fixed field, but also including meta data or schema. Also, big data is based on massive data and is needed of long term/strategic approach, and it is rarely required to instantly make decisions. The complexity of the process is high, the importance of unstructured data is high, and the processing and analyzing flexibility is high [10].

Analysis technique is divided into text mining, opinion mining, social network analysis, and clustering analysis [5].

2.2. Web Crawler

Web crawler is the function which is needed to collect information of web pages such as SNS. Web crawler is like a net or a spider web that gathers information from the webpage.

In the internet, to collect webpage delivery text or information of server as same as HEAD of client requirement to client. Using crawler while user does not enter into webpage, analyze it one by one and extracting wanted information automatically analyze web page and extract information [11].

When webpage is crawled, crawl and save meta data, HTML, title or context, image in database. When analyzing, remove overlapped page or spam article, document clustering. For this continuous process, crawler analyze webpage and deliver information to client.

2.3. Korean Morphological Analysis

Korean Morphological analysis is at early stage of natural language processing, taking apart sentences, analyzing words in morphologic view, and determining as a categorization or a part of speech. Because Korean word appeares in various context of language classification, result of morphological analysis is given possible nominee and analyzed to morpheme. Also, Korean has various function of verb of separate word, so while it is hard to effectively classify sentence, result of morphological analysis is expressed as morphological structure. During morphological analysis, extracting noun is an important function[12].

Korean Morphological analysis is classified as language dependent model, language Independent model, and etc. Language independent model is a two-level model applied to Korean and in two-level model, studying rule by machine learning, lexical transducer is suggested. The other ways are language subordinate way considering feature of Korean[13].

3. System Design

3.1. System Architecture

When selecting existing direction, commercial rights analysis or location information based on structured data was gained. But, structured data was limited and hard to quickly get information in live.

In this system, to improve this weakness, while SNS is active, information using structured data as well as unstructured data such as Big Data of SNS should be obtained.

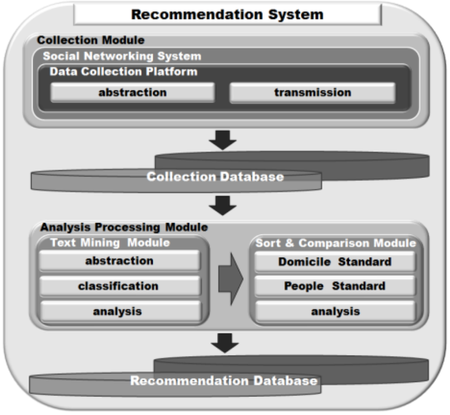


Figure 1. System Architecture

Figure 1 is a graphing system structure. This system is composed of Collection Module, Collection Database, Analysis Processing Module, Recommendation Database in Recommendation System. To get user’s writing of SNS, data must be collected in Collection Module. When collecting data, relevant SNS homepage should be crawled using the crawler.

The crawled data handles data and extracts the words to remove duplicating documents or for document clustering. This data is saved in Collection Database. The database used in this system uses My SQL. The Text Mining Module of Analysis Processing Module uses the hanannum morphological analyzer provided by KAIST to analyze morpheme with the postposition as the criteria. Sort & Comparison Module of Analysis Processing Module has the standard of postposition and is categorized as data analyzed users (sex, age), time, location (domicile, spot), and action.

Analyzed data is ranged by two criteria. One is user, another is domicile. In data ranged by user, it is known that user goes frequently which spot(Company Name). In data ranged by domicile, compare two ranged data following grasping frequency of address which user wants to get. For compared data, compare current population which is what type of users go with spot(Company Name) which is what spot those users go, and inform that which spot(Company Name) doesn’t exist in that.

3.2. System Flow Chart

Figure 2 shows an entire system flow. First, to select direction, get data from SNS base on unstructured data. For this, use crawler to crawl SNS. Get crawled data of hwp type and remove unnecessary context or overlapped documents, spam, etc. Following that, preprocessing context which SNS user write. Save preprocessed data in Collection Database.

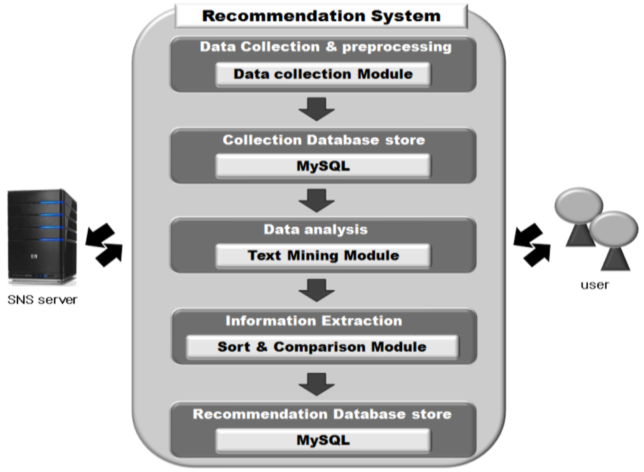


Figure 2. System Flow Chart

Bring saved data as Analysis Processing Module, and first, in Text Mining Module Morphological, analyze users’ talk. After, Morphological analyzed data is classified postposition based on already gotten criteria. Range by users and address by the classified data brought by Sort & Comparison Module. Compare those data of two types and extract information of direction. Finally, save it in Recommendation Database. Finally, save it in Recommendation Database.

4. System Implementation

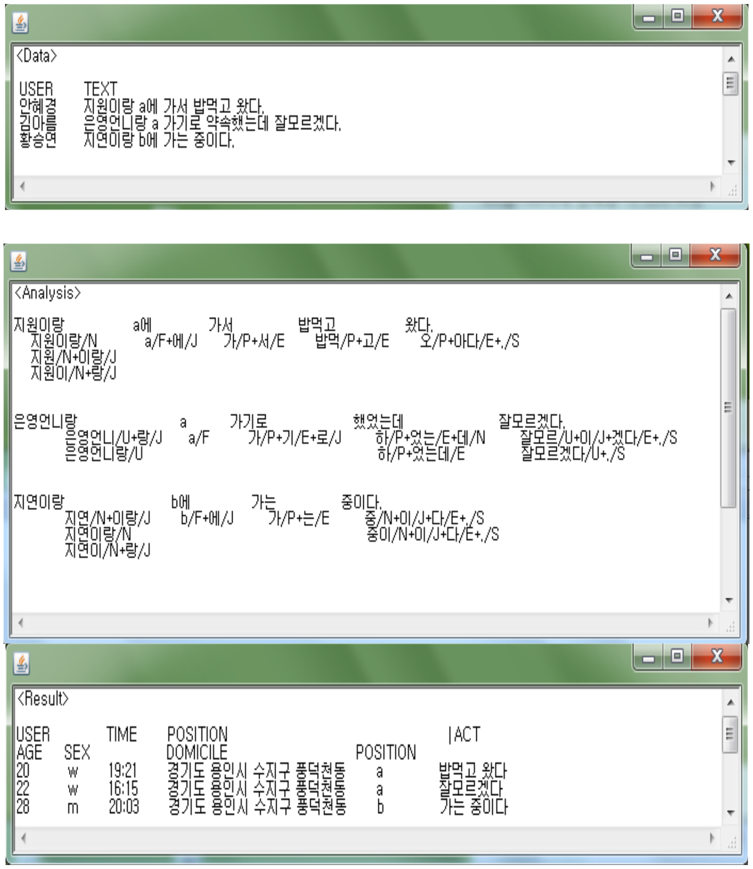
Table 1 shows system in test. The hardware specification is Intel Pentium 2.10GHz, 4GB RAM. Also, open source crawler and hannanum Morphological Analyzer is used, and using JAVA implement it as same as system flow factually.

Table 1. Experiment Environment

RAM	4GB
Programming Language	Java
Database	MySQL
CPU	Intel Pentium 2.10GHz

In test of prototype of this paper, location was selected as Pungdeokcheon-dong, Suji-gu, Yongin-si, Gyeonggi-do, Korea, and in these spot, extract information for Gyeonggy-do.

Figure 3 shows that after saving the writing of users in crawled data and put it into Morphological Analyzer. Below figure shows morphological analyzed data ranged by postposition. If the postposition in sentence is "*a*(아), *eui*(의), *kwa*(과), *eun*(은), *nun*(는), *lee*(이), *ka*(가), *rang*(랑), *hamkye*(함께)", it is classified to user. If it is "hour, minute, time," it is classified to time, if it is "*ae*(에), *aeseo*(에서), *eul*(을), *reul*(를)", it is classified to location, and if it is "*da*(다), *eo*(어)", it is classified to action. In Figure 3, a and b are spot name. Classified data is ranged by user and address, and draws an information.



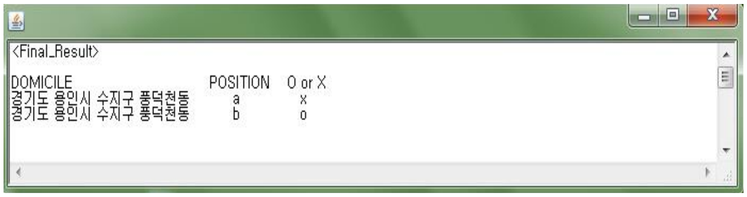
```
<Data>

USER      TEXT
안혜경    지원이랑 a에 가서 밥먹고 왔다.
김아름    은영언니랑 a 가기로 약속했는데 잘모르겠다.
황승연    지연이랑 b에 가는 중이다.
```

```
<Analysis>

지원이랑          a에          가서          밥먹고          왔다.
  지원이랑/N        a/F+에/J     가/P+서/E      밥먹/P+고/E      오/P+아다/E+./S
  지원/N+이랑/J
  지원이/N+랑/J


은영언니랑              a       가기로            했었는데              잘모르겠다.
        은영언니/U+랑/J   a/F      가/P+기/E+로/J      하/P+었는/E+데/N       잘모르/U+이/J+겠다/E+./S
        은영언니랑/U                                  하/P+었는데/E          잘모르겠다/U+./S


지연이랑               b에          가는           중이다.
        지연/N+이랑/J     b/F+에/J      가/P+는/E        중/N+이/J+다/E+./S
        지연이랑/N                                      중이/N+이/J+다/E+./S
        지연이/N+랑/J
```

```
<Result>

USER           TIME     POSITION                          |ACT
AGE    SEX              DOMICILE                  POSITION
20      w      19:21    경기도 용인시 수지구 풍덕천동      a       밥먹고 왔다
22      w      16:15    경기도 용인시 수지구 풍덕천동      a       잘모르겠다
28      m      20:03    경기도 용인시 수지구 풍덕천동      b       가는 중이다
```

Figure 3. Collection and Analysis Module

Figure 4 is a screen used to derive the total result. The result of the experiment implementing this system came out as shown in the figure.



```
<Final_Result>

DOMICILE                          POSITION   O or X
경기도 용인시 수지구 풍덕천동         a          x
경기도 용인시 수지구 풍덕천동         b          o
```

Figure 4. Result of Implementation

Figure 4 shows that in the specific region of Suji-gu, Yongin-si, Gyeonggi-do, Korea when comparing a location which users frequently go with a and b, in result, b existed but a did not exist. For this test, deduction a which is popular in this spot, and draw informative data.

5. Performance Evaluation

In this paper, because accuracy is important, performance evaluation was implemented. For experimental condition, after extracting 30 SNS data which aimed specific spot and tested ten times in the same direction. Compare probability of success with probability of fail, and know whether this system successfully recommend direction.

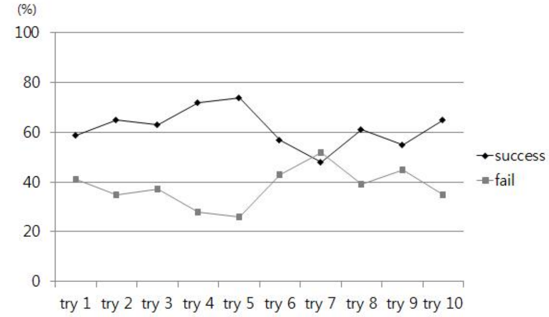


Figure 5. Result of Experiment

Figure 5 shows accuracy of this system. Performance evaluation using 300 data. In result of 10 times of each 30 SNS data, success rate of first performance is 69%, next performance is about 65%, next performance is about 63%, next performance is about 72%, next performance is about 74%, next performance is about 57%, next performance is about 48%, next performance is about 61%, next performance is about 55%, and the next performance is about 65%. Finally the average success rate is 61.9%, and the average failure rate is 38.1%.

6. Conclusion

In this paper, design and implement test with data which is Big Data of writing of SNS user to make it useful information. The result of this test performed 10 times showed a probability of 61.9% success rate.

In the future, increasing the probability of the success rate and increasing the reliability of the system is the good. Using Web sites as well as SNS, extract informative data, increase the accuracy, and widen target of spot.

Acknowledgements

This article is a revised and expanded version of a paper entitled [A Study on the Effective Location Recommendation] presented at International Symposium on Advanced and Applied Convergence held on November 14-16, 2013 at Seoul, Korea.

References

[1] S. J. Shin, "SNS using Big Data Utilization Research," Journal of the Institute of Internet, Broadcasting and Communication (JIIBC), vol. 12, no. 6, **(2012)**, pp. 267-272.

[2] K. Y. Lee, J. J. Kang, H. K. Ahn, K.H. Kim, G. S. Choi, S. J. Choi and S. J. Oh, "Location Recommendation System using Big Data," Advanced and Applied Convergence Letters (AACL), vol. 1, **(2013)**, pp. 9-10.

[3] N. J. Li, K. H. Choi, K. S. Jang and D. R. Shin, " RFID-based Secure Communication for Smart Device in Future Home Network Environment," International Journal of Internet, Broadcasting and Communication (IJIBC), vol. 5, no. 1, **(2013)**, pp. 18-22.

[4] J.J. Kim, J.J. Kang, K.Y. Lee, "Recovery Methods in Main Memory DBMS," International Journal of Advanced Smart Convergence (IJASC), vol. 1, no. 2, **(2012)**, pp. 26-29.

[5] M. M. Kang, S. R. Kim and S. M. Park, "Analysis and Utilization of Big Data," Korean Journal of the Institute of Information Scientists and Engineers (KJIISE), vol. 30, no. 6, **(2012)**, pp. 25-32.

[6] H. N. Kim, "Trend and Implication of Big Data," Korea Information Society Development Institute (KISDI), vol. 24, no. 19, **(2012)**, pp. 49-67.

[7] J. S. Kim, "Big Data Utilization and Related Technique and Technology Analysis," Korean Journal of the Contents Association (KJCA), vol. 10, no. 1, **(2012)**, pp. 34-40.

[8] G. Vishal and S. L. Gurpreet, "A Survey of Text Mining Techniques and Applications," Journal of Emerging Technologies in Web Intelligence (JETWI), vol. 1, no. 1, **(2009)**, pp. 60-76.

[9] McKinsey, "Big Data : The Next Frontier for Innovation, Competition, and Productivity, McKinsey and Company", (**2011**).

[10] S. W. Jo, "The Technology of Big Data Period", KT Advanced Institute of Technology Central Institute, **(2011)**.

[11] D. J. Kim, "Subject Information Service based on Web Crawler", Master's Thesis. Inha University Gradute School of Engineering, **(2008)**.

[12] J. D. Kim, H. C. Rim, J. D. Park and J. S. Lee, "Evaluation Method for Korean Morphological Analysis System and it's Application to MATEC99," The Conference of Special Interest Group in Human Language Technology (SIGHT), **(1999)**, pp. 44-49.

[13] S. S. Kang, "Korean Morphological Characteristics and Morphological Analysis," Korean Journal of the Institute of Information Scientists and Engineers (KJIISE), vol. 64, no. 8, **(1994)**, pp. 47-59.

Authors

Ki-Young Lee, h received his B.S. degree in Computer Science at Soongsil University in 1984. In 1988 and 2005, he received M.S. and Ph.D. degrees in Databases at Konkuk University, respectively. From 1984 until 1991, he worked for Korea Institute of Ocean Science and Technology (KIOST) as a researcher in Data Information and Processing department. He is currently a professor at the department of Medical IT and Marketing at Eulji University. He is also the head of department of S/W development in Bio-Meditech Regional Innovation Center at Eulji University. He is the director of the Korea Institute of Internet, Broadcasting & Communication(IIBC), and the director of the Korea Electronics Engineers(IEEK). His research interests include spatial databases, geographic information systems (GIS), location-based services (LBS), u-Healthcare, ubiquitous sensor network (USN), moving objects databases, and telematics etc.

Jeong-Jin Kang, he is currently the faculty of the Department of Information and Communication at Dong Seoul University in SeongNam, Korea since 1991, and currently the President of the Korea Institute of Internet, Broadcasting and Communication (IIBC). During 3 years from Feb. 2007 to Feb. 2010, he worked as a Visiting Professor at the Department of Electrical and Computer Engineering, The Michigan State University. He was a lecturer of the Department of Electronic Engineering at (Under) Graduate School (1991-2005), The Konkuk University. Dr. Kang is a member of the IEEE Antennas and Propagation Society(IEEE AP-S), the IEEE Microwave Theory and Techniques Society (IEEE MTT-S), and a member of the Korea Institute of Internet, Broadcasting and Communication (IIBC), Korea. His research interests involve Smart Mobile Electronics, RF Mobile Communication, Smart Convergence of Science and Technology, RFID/USN, u-Healthcare and ultrafast microwave photonics, as well as GIS, LBS, moving objects databases, and telematics etc.

Hye-Kyoung Ahn, she is currently a student in the department of Medical IT and Marketing at Eulji University. Her research interests include artificial intelligence(AI), search engine, text mining, Big Data, spatio-temporal database, etc.

Kyu-Ho Kim, he received B.S., M.S., and Ph.D. in Computer Science at Kwangwoon University in 1989, 1991, and 1998, respectively. From 1992 until 2007, he was a professor at the department of Internet Information at Seoul Health College. He is currently a professor at the department of Medical IT and Marketing at Eulji University. Since 2001, he is serving as the director of Korea Society of Computer Information (KSCI). His research interests include network management, ubiquitous sensor networks, and U-healthcare etc.

Gyoo-Seok Choi, he received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Yonsei University, Seoul Korea, in 1982, 1987, and 1997, respectively. He worked at the laboratory of DACOM Company as a researcher from 1987 to 1990. He also worked at the laboratory of SK Telecom Company as a senior researcher from 1991 to 1996. He is currently a professor at the Dept. of Computer Science in Chungwoon University. He is a vice-president of the Korea Institute of Internet, Broadcasting and Communication (IIBC). His current research interests include Artificial Intelligence, Telematics, Mobile Computing, etc.

Sung-Jai Choi, he was born in cheonan si, chung cheong nam – Do, Republic of Korea. He received the B.S. degree in electronics from Chungnam National University , Daejeon, Korea, and M.S. degree in electronics from Hanyang University , Seoul, Korea, and Ph.D. degrees in electronic engineering from MyongJi University, Yongin , Gyong gi - Do , 1n 1981 and 1985 and 2004 , respectively. Since 1988, he has been a professor at the department of Electronic Engineering of Gachon University with he’s main interests involve semiconductor manufacturing technology, electronic circuits analysis and fault testing , superconductor , Smart Mobile Electronics, RF Mobile Communication , Smart Convergence of Science and Technology , RFID/USN AND U-Healthcare , as well as New Media Service. Professor Choi is a Member of the Korea Institute of Internet, Broadcasting and Communication (IIBC), Korea , The Korean Information Processing Society (KIPS) , The Institute of Electronics Engineers of Korea (IEEK) and Korean Institute of Electrical and Electronic Material Engineering (KIEEME)

Sun-Jin Oh, he received the Bachelor of Engineering degree at Han Yang University, Korea, in 1981, and the Post Bachelor and the Master of Science degree in Computer Science at Wayne State University and University of Detroit Mercy, U.S.A. in 1987 and 1989 respectively. He completed the Ph.D. course in Computer Science at Oklahoma State University, U.S.A. and received the Ph.D. degree in Computer Science at Catholic University DG in 1999. He received the MBA degree in the Graduate School of Business Administration at Ajou University recently. Currently, he is the professor at the department of Computer and Information Science in Semyung University, Korea. He is the director of the Korea Institute of Internet, Broadcasting & Communication (IIBC), His research interests include Wireless Mobile Ad-Hoc Network, Ubiquitous Sensor Network, Location and Mobility Management in Mobile Computing, Mobile Games, Distributed Multimedia Systems, Cloud Computing and Big Data Analysis. He is a member of the IASTED, ACIS, IIBC, KIPS, KIIS and KMMS.