



텍스트마이닝을 활용한 연구동향 분석: 소셜네트워크서비스를 중심으로

윤혜진¹ · 김창식² · 광기영^{3*}

¹배화여자대학교 글로벌관광과

²배화여자대학교 글로벌관광과/국민대학교 비즈니스IT전문대학원

³국민대학교 경영대학/비즈니스IT전문대학원

Research Trends Investigation Using Text Mining Techniques: Focusing on Social Network Services

Hyejin Yoon¹ · Chang-Sik Kim^{2*} · Kee-Young Kwahk³

¹Department of Global Tourism, Baewha Women's University, Seoul 03039, Korea

²Department of Global Tourism, Baewha Women's University, Seoul 03039, Korea/Graduate School of Business IT, Kookmin University, Seoul 02707, Korea

³College of Business Administration/Graduate School of Business IT, Kookmin University, Seoul 02707, Korea

[요 약]

본 연구의 목적은 소셜네트워크서비스 주제에 관한 연구동향을 조사하는 것이다. 연구의 목적을 달성하기 위해서 웹오브사이언스 데이터베이스에서 제목에 'Social Network Service(SNS)'를 포함하는 1994년부터 2016년까지 출판된 논문 초록 308편을 분석하였다. 본 연구에서는 텍스트마이닝 기법 중에서 최근 많이 적용되는 토픽모델링기법을 활용하였다. 토픽모델링 분석결과 20개의 토픽(신뢰, 지지, 만족 모델, 조직 지배구조, 모바일 시스템, 인터넷 마케팅, 대학생 효과, 의견 확산, 고객, 정보보호, 건강관리, 웹 협업, 방법, 학습 효과, 지식, 개인 이론, 아동 지지, 알고리즘, 미디어 참여, 문맥 시스템)이 도출되었다. 또한 시계열회귀분석 결과 모든 토픽은 상승 추세로 나타났다.

[Abstract]

The objective of this study was to examine the trends on social network services. The abstracts of 308 articles were extracted from web of science database published between 1994 and 2016. Time series analysis and topic modeling of text mining were implemented. The topic modeling results showed that the research topics were mainly 20 topics: trust, support, satisfaction model, organization governance, mobile system, internet marketing, college student effect, opinion diffusion, customer, information privacy, health care, web collaboration, method, learning effectiveness, knowledge, individual theory, child support, algorithm, media participation, and context system. The time series regression results indicated that trust, support satisfaction model, and remains of the topics were hot topics. This study also provided suggestions for future research.

색인어 : 소셜네트워크서비스, 연구동향, 텍스트마이닝, 토픽모델링, 시계열분석

Key word : Social Network Services, Research Trends, Text Mining, Topic Modeling, Time Series Analysis

<http://dx.doi.org/10.9728/dcs.2018.19.3.513>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 23 December 2017; Revised 07 January 2018

Accepted 25 March 2018

*Corresponding Author; Chang-Sik Kim

Tel: +82-10-4761-3168

E-mail: solo21solo@naver.com

I . Introduction

Social network services (SNSs) have substantially changed our lives and how we interact with one another. Recent research shows that more and more people are using SNSs such as Facebook and Twitter for socializing with friends, sharing information, and other reasons [1], [2]. Many companies are also using social media to interact with their customers and provide a variety of information and services [3]. Today, SNSs serve as a significant tool facilitating communication and organizational activities among friends, customers, and business, which has led to increase of research article on social media [4]. Previous studies have generally focused on the causal-chain framework to explain the causality between research constructs.

However, the phenomenon of SNSs still remain new to academia. Although the proliferation of SNSs uses, there are lack of research in a wide range of research topics [5]. Thus, researchers, marketers, and decision makers should be well prepared to overcome the challenges and find opportunities from the emerging phenomenon of SNSs in a wide variety of fields. Ngai et al. (2015, p. 34) also suggest that “more studies on social media will be conducted and further results will be available in the coming years” [5]. To find hidden patterns, trends, and opportunities for future, text mining is regarded as a suitable methodology [2]. Text mining, an extension of data mining, can extract useful knowledge from overloading text documents [2].

In this context, this paper investigates the research trends of academic articles in SNSs from 1994 to 2016 using text mining techniques. This study provides a comprehensive overview on the SNSs related literature, and identifies the main research topics and themes. It also highlights opportunities for future research. The remainder of this article is organized as follows. Section 2 is a brief literature review of text mining and social network services. Section 3 describes the methodological approach used to collect and analyze the data. Then, the key findings are discussed in Section 4. Lastly, conclusions are summarized in Section 5 with suggestions for future research directions.

II . Literature Review

2-1 Text mining

Text mining is an emerging technique that aims to extract useful information from unstructured text documents [2]. This

automated technique is a systematically extract, manage, integrate, exploit, and identify knowledge from texts [2], [6], [7]. Many researchers have successfully used the application of text mining for finding meaningful trends, patterns, models, or rules from unstructured textual data (e.g., text files, chat messages, and email) in a variety of fields, including business [2], education [7], health [8]; then creating interpretation that describes interesting trends and patterns in the text [2], [8].

Text mining has largely used for analyzing the literature and identifying research trends [6]. For example, the study of Moro, Cortez, and Rita (2015) analyzed the published articles in business intelligence for the banking industry from 2002 to 2013 using text mining [9]. They found meaningful trends that customer relationship management has become significant in the banking industry, thus suggested directions for future research. Hung (2012) also analyzed the research trends of e-learning literature from 2000 to 2008 using text mining; suggested that e-learning applications in medical education and training are growing areas for future research [7].

2-2 Social Network Services

Ji, Hwangbo, Yi, Rau, Fang, and Ling (2010, p. 1104) defined Social Networking Services (SNSs) as “SNS is a web-based service that forms relationships between individuals by providing profile bases which include individual information, makes social interaction between participants easy by providing users with functions to communicate with each other, and provides a platform for users to share information and contents” [10]. Similarly, Boyd and Ellison (2007 p. 211) illustrated that SNSs allow “individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system” [11].

A SNS, one of the newest types of digital media, is emerging worldwide as a communication tool. According to an eMarketer's report, approximately one in four people used SNSs globally in 2013; the number of SNS users increased from 1.47 billion in 2012 to 1.73 billion in 2013 and is predicted to rise up to 2.55 billion by 2017 [12]. This trend indicates that an SNS is a powerful platform from which people's thoughts and social interactions are shared cross-nationally; which influences our daily lives.

Social scientists are beginning to explore what impact SNSs could have on current society. They have opened up a broader discussion on the concepts and trends of SNSs [11], SNSs' benefits [13], social capital [14], privacy issues [15], and teenage use [16]. Previous studies have usually conducted to empirically

investigate the relationships among antecedents, mediators, moderators, and outcomes regarding the perceptions and/or behaviors of social media users. These studies have shown the potential benefits of SNSs as a marketing tool for a wide range of business and also a communication. Yet, many critical areas—for example, impacts of social media, social power, and cultural differences—have not received research attention [5]. Curiously, despite the rise of social media research, few have attempted to address the research trends of SNSs literature.

III. Methods

3-1 Target of Analysis

This study utilized abstracts as data sources for topic modeling. Excel and SAS Enterprise Miner 14.1 (SAS) were employed for the pre-processing of the unstructured data from the database. All the research papers from the 219 journals published for the last 23 years, from 1994 to 2016, were retrieved through the online databases of web of science. A total of 308 papers contained the term ‘Social Network Services (SNSs)’ in the titles were collected and the abstracts for the papers constituted a database for the study. Table 1 shows the number of articles for each journal in the study period and Figure 1 shows the articles trends. The first seven SNSs articles were published before 2000. From 2010, the annual number of publications on SNSs rapidly increased in the past decades.

3-2 Process of Analysis

Topic modeling in SAS is performed in 4 steps: 1) Text Import, 2) Text Parsing, 3) Text Filter, and 4) Topic Modeling. In Text Import, text data are imported. In Text Parsing, terms contained in the imported documents are parsed in order to quantify the data. In Text Filter, unnecessary terms are eliminated. In Text Topic, a list of relevant terms is created to allow users to easily pick topics that accurately represent a main theme or idea (i.e., getting started with SAS Text miner 14.1).

After topic modeling, this study performs a time series regression analysis using SPSS to analyze the research trends of the last 23 years. The time series regression has year as an independent variable and 20 different topics as dependent variables.

표 1. 소셜네트워크서비스 논문 기간별 트렌드

Table 1. Articles contained ‘SNS’ in the Title

Periods	94~99	00~05	06~11	12~16	Total
Article	7	11	71	219	308

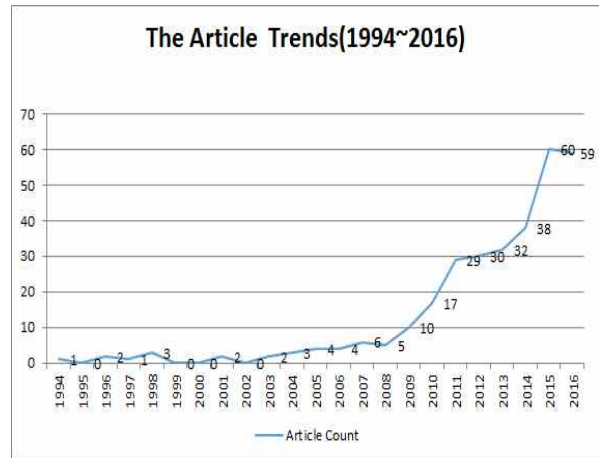


그림 1. 소셜네트워크서비스 논문 연도별 트렌드

Fig. 1. Article Trends contained ‘SNS’ in the Title

IV. Results

4-1 Topic Modeling

The name of each topic was defined through interpretation of researchers, based on the topic words. The main research topics detected from the analysis included friend, mobile system, organization governance, support, and web collaboration. The defined 20 topics for SNSs research was used for the further analysis of the study. Table 2 describes the extracted 20 main topics on SNSs research from 1994 to 2016, and lists the topic words which impose higher probability than other words in respective topics.

4-2 Time Series Analysis

This study performs a time series regression analysis using SPSS. The time series regression has year as an independent variable and 20 different topics as dependent variables. A research topic that has a significant parameter estimate, at an alpha level of 0.05, with a positive coefficient is called a ‘hot’ topic, and a topic that has a significant parameter estimate with a negative coefficient is called a ‘cold’ topic. A hot topic indicates a rising trend, while cold topic indicates a falling trend [17], [18].

As a result, twenty hot topics on the SNSs research were derived ($p < 0.05$), as shown in Table 3 and Figure 2. The hot topics were trust, support, satisfaction, organization management, mobile device, internet marketing, college student, opinion diffusion, customer, information privacy, health care, web collaboration, friend, learning effectiveness, knowledge, capital, child support, algorithm, media participation, and context system.

표 2. 소셜네트워크서비스 논문 토픽모델링 결과

Table 2. Results of Topic Modeling (1994~2016)

Topic	Keyword	Freq
T13	friend	friend, method, people, difference, user
T05	mobile device	device, mobile, system, application, message
T04	organization management	organization, management, governance, sector, ecosystem
T02	support	support, member, friend, client, health
T12	web collaboration	web, collaboration, interaction, discovery, reputation
T16	capital	capital, individual, theory, framework, intervention
T15	knowledge	knowledge, answer, question, learning, tie
T11	health care	care, health, patient, hospital, utilization
T09	customer	customer, consumer, company, firm, author
T03	satisfaction	intention, continuance, satisfaction, enjoyment, model
T10	information privacy	privacy, benefit, information, user, protection
T20	context system	context, system, recommendation, relation
T14	learning effectiveness	customer, learning, effectiveness, method, opinion
T17	child support	parent, youth, child, support, strain
T18	algorithm	algorithm, performance, search, location, method
T19	media participation	media, participation, customer, public, challenge
T01	trust	trust, selection, path, recommendation, algorithm
T08	opinion diffusion	node, diffusion, opinion, relay, simulation
T07	college student	student, college, effect, variable, university
T06	internet marketing	library, internet, marketing, science, view

V. Conclusion

The aim of this study was to investigate the trends on SNSs. The current research identified primary research themes in SNSs literature by analyzing 308 abstracts of the research articles from the web of science database published from 1994 to 2016. Topic modeling and time series regression techniques were employed for the analysis. The findings of this study are as follows.

1. Topic modeling was applied to determine the SNSs topics which have been researched most by scholars. Twenty major research topics were identified: including friend, mobile device, organization management, support, capital, web collaboration, knowledge, health care, customer, satisfaction, information privacy, context system, algorithm, media participation, trust, opinion diffusion, college student, and internet marketing.

2. Yearly trends of the identified research topics were calculated to see the advance or decline of each topic as time elapsed. The analysis of a time-serial linear regression showed that all twenty topics were hot topics in the SNSs literature.

표 3. 소셜네트워크서비스 논문 시계열회귀분석 결과

Table 3. Results of Time Series Regression (1994~2016)

Topic	regression coefficient	p-value	Trends	
T01	trust	.201	.001	Hot
T02	support	.139	.006	Hot
T03	satisfaction	.242	.000	Hot
T04	organization management	.273	.000	Hot
T05	mobile device	.341	.000	Hot
T06	internet marketing	.099	.003	Hot
T07	college student	.154	.000	Hot
T08	opinion diffusion	.181	.002	Hot
T09	customer	.241	.000	Hot
T10	information privacy	.237	.000	Hot
T11	health care	.139	.000	Hot
T12	web collaboration	.298	.000	Hot
T13	friend	.300	.000	Hot
T14	learning effectiveness	.202	.000	Hot
T15	knowledge	.263	.000	Hot
T16	capital	.254	.000	Hot
T17	child support	.138	.001	Hot
T18	algorithm	.219	.000	Hot
T19	media participation	.203	.000	Hot
T20	context system	.191	.000	Hot

The results indicated that existing research on SNSs covers a wide range of topics. The topics of SNSs research will continue to increase in the next few years. Wilson, Gosling and Graha (2012) also pointed out that the research perspectives, questions, and methods in SNSs literature are quite diverse and fragmented [19]. Despite such difficulties, the findings of this study provide a comprehensive overview for scholars interested in SNSs research.

The finding suggests topics needing additional research. Existing studies on SNSs generally rely on the topics of friend, mobile device, support, capital, web collaboration, knowledge, health care, customer, and satisfaction. SNSs can play a key role in facilitating and encouraging online communication among friends, customers or between customers and organizations. Researchers could raise significant questions regarding how online interaction on SNSs influences individuals' social network, customers' satisfaction, or organization management. In this context, social capital can be a popular theoretical framework [20]. However, there is a long debate whether social capital derives from the network of strong relationships or from the weak relationships in sociology [20]. Thus, future researchers should pay more attention to the network structure of social capital on SNSs. Additionally, as shown in the Table 1 and 2, scholars have taken a variety of topics and approaches to investigate the popular communication tool. They have generally focused on the impact of SNSs on the ways in which networks are established, maintained, and mediated. Yet, more research is needed to clarify the broader scope of SNSs in the current literature.

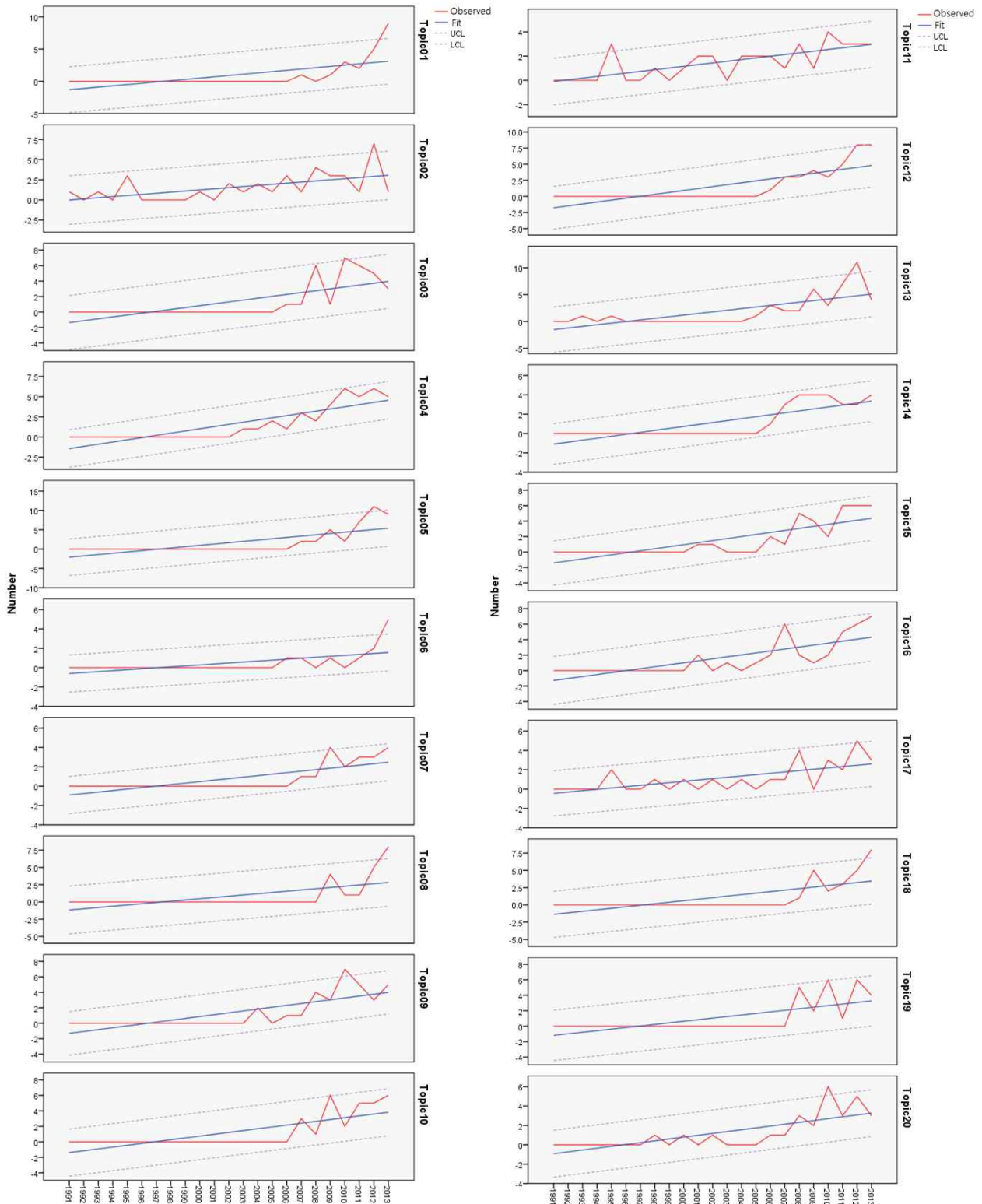


그림 2. 소셜네트워크서비스 논문 각 토픽별-연도별 트렌드
 Fig. 2. Results of Topic Trends

The findings highlight that future research is more needed to address the growing market of SNSs and its application in a wide range of areas. However, this paper selected abstracts as the primary data source, and the quality of abstracts could have affected the research findings. Thus, the findings of this study may not reflect any significant information.

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2015S1A3A2046711).

References

- [1] A. M. Kaplan and M. Haenlein, "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business Horizons*, Vol. 53, No. 1, pp. 59-68, 2010.
- [2] W. He, W. S. Zha and L. Li, "Social Media Competitive Analysis and Text Mining: A Case Study in the Pizza Industry," *International Journal of Information Management*, Vol. 33, No. 3, pp. 464-472, 2013.
- [3] V. M. Sinderen and J. P. A. Almeida, "Empowering Enterprises through Next-generation Enterprise Computing," *Enterprise Information System*, Vol. 5, No. 1, pp. 1-8, 2011.
- [4] R. Hanna, A. Rohm and V. L. Crittenden, "We're All Connected: The Power of the Social Media Ecosystem," *Business Horizons*, Vol. 54, No. 3, pp. 265-273, 2011.
- [5] E. W. Ngai, S. S. Tao and K. K. Moon, "Social Media Research: Theories, Constructs, and Conceptual Frameworks," *International Journal of Information Management*, Vol. 35, No. 1, pp. 33-44, 2015.
- [6] D. Delen and M. D. Crossland, "Seeding the Survey and Analysis of Research Literature with Text Mining," *Expert Systems With Applications*, Vol. 34, pp. 1707-1720, 2008.
- [7] J. L. Hung, "Trends of E-learning Research from 2000 to 2008: Use of Text Mining and Bibliometrics," *British Journal of Educational Technology*, Vol. 43, No. 1, pp. 5-16, 2012.
- [8] L. Li, R. L. Ge, S. M. Zhou and R. Valerdi, "Guest Editorial Integrated Healthcare Information Systems," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 16, No. 4, pp. 515-517, 2012.
- [9] S. Moro, P. Cortez, and P. Rita, "Business Intelligence in Banking: A Literature Analysis from 2002 to 2013 using Text Mining and Latent Dirichlet Allocation," *Expert Systems with Applications*, Vol. 42, No. 3, pp. 1314-1324, 2015.
- [10] Y. G. Ji, H. Hwangbo, J. S. Yi, P. P. Rau, X. Fang and C. Ling, "The Influence of Cultural Differences on the Use of Social Network Services and the Formation of Social Capital," *Intl. Journal of Human -Computer Interaction*, Vol. 26, No. 11-12, pp. 1100-1121, 2010.
- [11] D. M. Boyd and N. B. Ellison, N. B., "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, Vol. 13, No. 1, pp. 210-230, 2007.
- [12] eMarketer, "Social Networking Reaches nearly One in Four around the World," *eMarketer*, Retrieved from <http://www.emarketer.com/Article/Social-Networking-Reach-es-Nearly-One-Four-Around-World/1009976>, June 18, 2013.
- [13] S. Utz, "The (Potential) Benefits of Campaigning via Social Network Sites," *Journal of Computer-Mediated Communication*, Vol. 14, No. 2, pp. 221-243, 2009.
- [14] T. A. Pempek, Y. A. Yermolayeva and S. L. Calvert, "College Students' Social Networking Experiences on Facebook," *Journal of Applied Developmental Psychology*, Vol. 30, No. 3, pp. 227-238, 2009.
- [15] D. J. Houghton and A. N. Joinson, "Privacy, Social Network Sites, and Social Relations," *Journal of Technology in Human Services*, Vol. 28, No. 1-2, pp. 74-94, 2010.
- [16] U. Pfeil, R. Arjan and P. Zaphiris, "Age Differences in Online Social Networking: A Study of User Profiles and the Social Capital divide among Teenagers and Older Users in MySpace," *Computers in Human Behavior*, Vol. 25, No. 3, pp. 643-654, 2009.
- [17] T. L. Griffiths and M. Steyvers, "Finding Scientific Topics," in *Proceedings of the National academy of Sciences*, Vol. 101, No. 2, pp. 5228-5235, April 2004.
- [18] C. -S. Kim, S. J. Choi and K. -Y. Kwahk, "Investigation of Research Trend in Information Systems Domain using Topic Modeling and Time Series Regression Analysis," *Journal of Digital Contents Society*, Vol. 18, No. 1, pp. 151-159, 2017.
- [19] R. E. Wilson, S. D. Gosling and L. T. Graham, "A Review of Facebook Research in the Social Sciences," *Psychological Science*, Vol. 7, No. 3, pp. 203-220, 2012.
- [20] Y. Zhang and L. Leung. "A Review of Social Networking Service (SNS) Research in Communication Journals from 2006 to 2011," *New Media & Society*, Vol. 17, No. 7 pp. 1007-1024, 2015.



윤혜진 (Hyejin Yoon)

2004년 : 중앙대학교 행정학사, 광고홍보학사, 문학사
2006년 : 경희대학교 호텔관광학과 관광학석사
2010년 : 경희대학교 호텔관광학과 관광학박사
2015년 : Indiana University (Bloomington) 여가행동학박사 (노인여가 전공)

2016년: 경희대학교 호텔관광대학 BK21 플러스 연구박사
2017년~현 재: 배화여자대학교 글로벌관광과 조교수
※ 관심분야 : 관광/여가 행동(Tourism/Leisure Behavior), 삶의 질(Quality of Life), 노인 여가(Older Adults' Leisure), 에이징 웰(Aging Well)



김창식 (Chang-Sik Kim)

2002년 : 경희대학교 산업정보대학원 (경영학석사)
2013년 : 국민대학교 비즈니스IT전문대학원 (경영정보학박사-비즈니스IT)

1995년~2014년: 한화호텔앤드리조트/한화에스앤씨
2015년~2018년: 국민대학교 비즈니스IT전문대학원 BK21 플러스 사업팀 계약교수
2018년~현 재: 배화여자대학교 글로벌관광과 조교수
※ 관심분야 : 환대관광정보기술(Tourism and Hospitality Information Technology), 텍스트마이닝(Text Mining), 머신러닝(Machine Learning), 딥러닝(Deep Learning), 데이터 애널리틱스(Data Analytics)



곽기영 (Kee-Young Kwahk)

1988년 : 서울대학교 경영대학 (경영학사)
1990년 : 한국과학기술원 경영과학과 (경영학석사)
1999년 : 한국과학기술원 테크노경영대학원 (경영정보학박사)

1990년~2002년: 삼성전자/삼성SDS
2003년~2005년: 계명대학교 경영대학
2006년~현 재: 국민대학교 경영대학/비즈니스IT전문대학원 교수
※ 관심분야 : 소셜네트워크분석 및 응용(Social Network Analysis and Its Application), 데이터 애널리틱스(Data Analytics), 지식경영(Knowledge Management) 등