http://dx.doi.org/10.5392/JKCA.2014.14.09.567

소셜 네트워크 서비스의 연구경향 분석: Twitter관련 연구 중심

Analysis of Research Trends on Social Network Service: focusing on the Studies of Twitter

하일규
영남대학교 컴퓨터공학과

Ilkyu Ha(ilkyuha@ynu.ac.kr)

요약

최근 들어, 소셜 네트워크 서비스(Social Network Service)의 도입과 함께 이를 다양한 목적으로 활용하고자 하는 연구가 활발하게 진행되고 있다. 소셜 네트워크 서비스 활용을 위한 연구를 진행하기 위해서는 우선 관련 문헌을 검토하고 연구 경향을 파악하는 것이 필요하다. 하지만, 소셜 네트워크 서비스가 보편화된 2007년 이후부터 진행된 관련 연구는 그 양이 방대하여 많은 양의 관련 연구 문헌을 조사하는 것은 상당히 어려운 작업이다. 따라서 본 연구에서는 소셜 네트워크 서비스 중 트위터를 중심으로 관련 연구들을 체계적으로 분석하여 연구의 경향성을 밝힌다. 특히 체계적인 문헌 조사와 분석을 위해 SLR(Systematic Literature Review) 기법을 이용한다. 문헌 조사를 위해 5개의 잘 알려진 문헌 제공처를 선정하고, 128개의 문헌 조사 대상 연구를 선정하였다. 다양한 경향성을 파악하기 위하여 2007년 이후 연구들의 그룹과 2011년 이후 연구들의 그룹으로 나누어 조사를 실시하였다. 조사 결과 2007년 이후 트위터 관련 연구는 “Application”, “User Content Analysis” 및 “User Activity” 주요 연구주제와 관련한 연구가 가장 많이 이루어지고 있었다. 이외에도 주요 연구주제별 세부 연구주제의 경향, 피인용수에 따른 연구 경향, 실험자료의 규모에 따른 연구경향 및 저자의 특성 등 다양한 관점에서 결과를 분석한다.

■ 중심어 : | 소셜 네트워크 서비스 | 체계적 문헌검토 | 트위터 | 트위터 분석 | 연구 경향 |

Abstract

Recently, with the introduction of social network services, studies that try to make use of them for the various purposes have been actively developed. In order to proceed with the research that takes advantage of social network services, it is necessary to review the relevant literature and to identify trends in researches. However, the researches of social network service since 2007 are massive amount, so to review the huge amount of relevant research literature is a very difficult task. Therefore, in this study, we analyze systematically the tendency of research related to social network service focusing on Twitter. Especially, we use the SLR(Systematic Literature Review) technique for systematic literature survey and analysis. For the literature survey, we select five well-known literature resource sites and 128 studies of literature that are surveyed. In order to identify various research trends, we conduct the survey with two research groups: researches since 2007 and researches since 2011. As a result of the investigation, since 2007, the researches associated with “Application”, “User Activity” and “User Content Analysis” main study topics have been mostly carried out. In addition to the result, the trend of secondary study topics in a main study topic, trends in research based on the number of citations and the scale of the experimental data and characteristics of the author are analyzed from a variety of perspectives.

■ keyword : | Social Network Service | Systematic Literature Review | Twitter | Twitter Analysis | Research Trend |

접수일자 : 2014년 07월 31일
수정일자 : 2014년 08월 11일
심사완료일 : 2014년 08월 12일
교신저자 : 하일규, e-mail : ilkyuha@ynu.ac.kr

I. 서 론

다양한 소셜 네트워크 서비스(Social Network Service)가 도입되면서 일상생활에서는 없어서는 안 될 중요한 의사소통 도구로 사용되고 있다. 소셜 네트워크 서비스는 "사용자 간의 자유로운 의사 소통과 정보 공유, 그리고 인맥 확대 등을 통해 사회적 관계를 생성하고 강화시켜주는 온라인 플랫폼"으로 정의된다[1]. 이러한 소셜 네트워크 서비스는 편리하고 유용한 의사소통 수단이라는 이점 때문에 다양한 분야에서 이를 활용하고자 하는 노력이 이어지고 있다. 지금까지 정치[7], 경제[9], 교육[11], 언론[39], 보건[12], 기후[40] 등 다양한 분야에 관한 연구가 결실을 맺고 있으며, 보다 고도의 서비스를 위한 연구가 활발히 진행되고 있다. 소셜 네트워크 서비스의 활용과 현상에 관한 연구를 진행하기 위해서는 무엇보다도 선행 연구의 조사를 통해 연구의 동향과 수준을 파악하는 것이 필요하다. 그러나 소셜 네트워크 서비스가 보급된 이후로 수없이 진행된 연구 문헌들을 모두 조사하여 연구의 경향성을 파악하기는 매우 힘든 일이다. 한편, 소셜 네트워크 서비스의 한 종류인 트위터(Twitter)도 2006년 10월 소개된 이후로 여러 가지의 편리한 점 때문에 빠른 성장을 보이고 있다. 즉, 140 문자로 이루어진 단문을 상대방의 허락없이 접근하여 조회할 수 있고, 팔로윙(following)-팔로워(follower) 관계를 통해 트윗(tweet) 발송, 응답(reply), 리트윗(retweet), 멘션(mention) 등 다양한 액션을 통해 의사를 전달할 수 있다[2].

본 연구에서는 소셜 네트워크 서비스에 관련된 방대한 양의 연구 문헌을 조사하여 연구의 경향성(변화의 추이 및 특징)을 밝히고자 한다. 특히 학계에서 많이 연구하고 있고, 서구에서 많이 이용되고 있는 소셜 네트워크 서비스인 트위터에 중점을 두어 관련 문헌을 조사하여 연구의 경향성을 파악한다. 보다 체계적인 문헌 조사를 위해 SLR(Systematic Literature Review) 기법[3]을 도입하여 여러 단계로 구성된 세부 절차에 의해 조사를 실시한다. 문헌 조사를 위해 학계에 잘 알려진 5개의 문헌 제공처를 선정하고, 각 문헌 제공처로부터 대표적인 문헌을 선별하는 과정을 거쳐 최종적으로 128개의 조사 대상 문헌을 선정한다. 연구의 경향성을 보다 자세하게 파악하기 위해 관련 연구를 2개의 그룹으로 나누어 조사한다. 즉, 트위터 서비스가 이루어진 직후인 2007년 이후부터 현재까지의 연구 문헌을 그룹 1로 하여 트위터 관련 연구의 전체적인 경향을 파악하고, 최근의 연구 경향을 다시한번 파악하기 위해 2011년 이후부터 현재까지의 연구문헌을 그룹 2로 하여 조사한다. 조사 대상 문헌으로부터 자료를 추출하기 전에 필요한 자료와 규모에 대한 계획을 세우고 계획에 따른 추출 작업이 실시된다. 수집된 자료는 가공과 처리를 거쳐 다양한 관점에서 분석된다. 분석 결과를 통해 연구의 경향성을 파악한다.

2장에서는 SLR기법과 트위터의 특성에 대하여 설명하고, 3장에서는 SLR기법에 의한 문헌 조사를 실시하며, 4장에서는 SLR기법에 의한 문헌 조사 결과를 분석하고 연구경향을 요약한다. 마지막으로 5장에서 결론을 맺는다.

II. 관련 연구

1. SLR(Systematic Literature Review) 기법

연구의 경향성을 보다 체계적으로 조사하기 위해 SLR 기법이 이용된다. SLR 기법은 문헌을 추출하고 분류하는데 사용되는 잘 알려진 방법이다. Kitchenham[3]는 SLR 기법을 특정한 연구 질문, 관심 있는 연구 분야 또는 현상에 대해 모든 관련 있는 문헌을 조사(identifying)하고 평가(evaluating)하며 해석(interpreting)하기 위한 방법으로 정의하고 있다. SLR은 초기에 3가지 주요 조사 단계로 구성되었다. 즉, 계획(planning), 실행(conducting) 그리고 보고(reporting) 단계이다. 그 후 조사 과정은 Lucas[4]에 의해 5 단계로 확장되었다. SLR은 다양한 문헌조사에 이용된다. 예를 들어 Lucas[4]는 소프트웨어공학분야 중 UML 관련 연구를 조사하고 분석하는데 사용하였으며, Antonescu[5]는 UML연구의 경향성을 파악하는데 이용하였다. 이러한 SLR 절차를 본 연구에서는 연구 주제를 중심으로 연구의 경향성을 파악하는 본 연구의 특성에 맞게 6 단계의

활동으로 확장하여 연구 조사에 이용한다. 기존의 절차에서 “Extraction Planning”단계와 “Analyzing”단계가 포함되었다. “Extraction Planning”단계는 추출할 항목을 결정하고 추출 규칙을 설정하는 단계이다. “Analyzing”단계는 추출된 데이터를 분석하고 경향성을 파악하는 단계이다. 6 단계는 [표 1]과 같다.

표 1. SLR의 절차

B.Kitchenham (2004)	F. Lucas (2009)	This paper (2014)	활동(Activities)
계획 (Planning)	Question Formularization	Question Formation	Setting up the purpose of the study and research questions for accurate data acquisition
실행 (Conducing)	Source Selection	Extraction Planning	Selection of main extraction items and rules for the extraction of data
	Studies Selection Process	Source Selection	Selection of literature resource sites that provides a wealth of research materials
	Information Extraction	Researches Selection	Selection of representative literatures from each resource site
	Extraction Execution	Data Extraction	Extraction of data items from the selected literatures
	–	Analyzing	Analysis of the extracted data, acquisition of the meaningful information, identifying trends
보고 (Reporting)	Reporting	Reporting	Publication of information and trends that have been grasped

2. 트위터의 특성

트위터는 사용자들 사이의 관심과 정보를 교환하기 위한 마이크로 블로깅 형태의 소셜 네트워크 서비스이다[6]. 트위터에서 주고받는 메시지는 트윗(tweet)이라는 140 문자의 짧은 텍스트로 되어 있으며, 텍스트 메시지뿐만 아니라 이미지나 비디오 등 멀티미디어 정보에 대한 링크도 포함되어 있다. 또한 해시태크(hashtag)나 ‘@’ 문자와 같은 다양한 축약된 표기법을 이용함으로써 단문이기는 하지만 매우 풍부한 내용을 포함할 수 있다. 트위터에서 사용자의 관계는 팔로잉(following)-팔로워(follower) 관계를 중심으로 이루어지며 자신이 팔로잉하고 있는 사용자로부터 트윗 메시지를 수신할 수 있으며, 자신을 팔로잉하고 있는 사용자(팔로워)에게 트윗 메시지를 전달할 수 있다. 메시지에 대해서는 응답(reply), 언급(mention) 및 리트윗(retweet)과 같은 기능으로 메시지에 반응할 수 있다. [그림 1]은 트위터의 기능을 요약한 것이다.

트위터에서는 다양한 API를 제공함으로써 다양한 형태의 연구를 가능하게 한다[39][40]. 따라서 트위터는 연구자들 사이에서 가장 인기있는 소셜 네트워크 서비스 중 하나이다. 즉, 과거의 트윗 내용을 조회하여 원하는 정보와 정보의 흐름을 파악하여 다양한 분야에 응용할 수 있다. 몇몇 연구에서는 이와 같은 API를 이용하여 지진, 허리케인과 같은 자연재해 발생에 대한 과거의 시간과 위치 정보를 파악하여 가공하고 이를 미래의 동일한 자연재해의 발생과 흐름을 예측하는데 이용한다[40]. 또는 대량의 일상 대화로 이루어진 트윗 데이터를 획득하여 그날의 핫이슈(hot issue), 토픽어(topic keyword) 그리고 감성(sentiment)과 같은 특별한 정보를 추출하기도 한다.

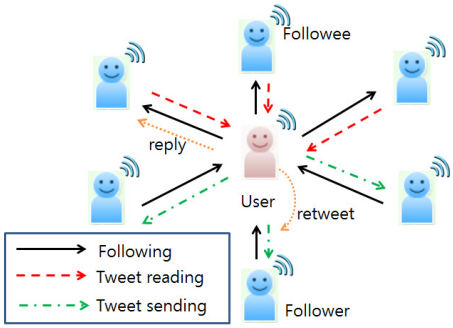

그림 1. 트위터의 활동들

한편, 트위터에 관련된 연구는 2007년 이후 빠르게 증가하고 있다. 동일한 키워드 “twitter” 또는 “tweet”

를 문헌의 제목에 포함하고 있는 자료를 주요 문헌 제공처에서 검색하여 보면 [그림 2]와 같이 매년 빠르게 증가하고 있는 것을 알 수 있다. 새로운 획기적인 의사소통 수단이 나타나지 않는 한 트위터의 사용과 연구는 계속될 것으로 예측된다.

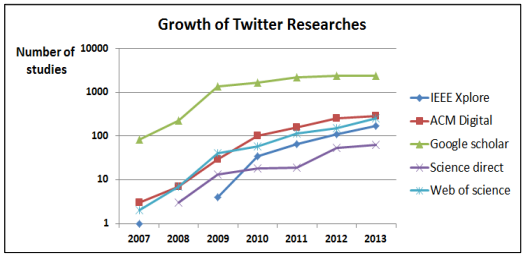

그림 2. 트위터 연구의 성장 (2014년 5월)

III. SLR기법에 의한 문헌 조사

SLR 기법에 의한 체계적 문헌 조사를 위해 본 연구에서는 [표 1]에서 제시한 6개의 단계에 따라 조사를 실시한다. 질의 형성(Question Formation) 단계, 데이터 추출계획(Extraction Planning) 단계, 문헌제공처 선정(Source Selection) 단계, 조사대상 문헌 선정(Researches Selection) 단계, 데이터 추출 단계(Data Extraction) 단계 그리고 마지막으로 분석(Analyzing) 단계이다. 각 단계에서 이루어지는 작업에 대하여 설명한다.

1. 질의 형성(Question Formation) 단계

질의 형성 단계는 정확한 데이터 추출과 효율적인 연구를 위하여 연구의 목적과 연구의 의문점을 설정하는 단계이다. 우선, 정확한 데이터 획득을 위해 문헌 조사에 앞서 연구의 경향성에 대한 몇 가지 의문점을 설정한다. 의문점을 통해 연구의 목적을 파악하고, 얻고자 하는 데이터의 형태와 종류를 파악한다. 아래와 같은 의문점을 설정한다.

(1) 소셜 네트워크 서비스인 트위터에서는 어떤 연구가 진행되고 있는가?
- 각 연구의 주요 연구 주제(main topic)는 무엇이고 어떻게 분류할 수 있는가?
- 각 연구의 주요 주제별 연구의 경향성과 특성은 어떠한가?

(2) 저자들의 특성은 무엇인가?
- 저자들의 전공분야, 소속국가, 저자의수, 출판연도 등은 어떤 경향성을 가지는가?

따라서 본 연구는 트위터 관련 연구 문헌 조사를 통해 주요 연구 주제를 파악하고 주제별 연구의 경향성과 특성을 파악함을 목적으로 한다. 아울러 문헌의 저자의 정보를 통해 트위터 관련 연구자의 특성을 파악한다.

2. 데이터추출 계획(Extraction Planning) 단계

데이터 추출 계획 단계에서는 데이터추출을 위한 주요 규칙과 추출 항목이 선정된다. 다음과 같은 데이터가 추출된다. 즉, 조사 대상 문헌의 제목, 출판년도, 저자명, 문헌의 종류, 주요 연구 주제, 세부 연구 주제, 문헌제공처의 인용횟수, 실험에 필요한 트윗 건수, 실험에 필요한 사용자 계정 수, 실험에 필요한 트윗 샘플링 기간, 저자의 전공, 저자의 소속국가, 저자의 소속기관, 저자의 수 등이다. [표 2]는 데이터추출항목과 용도를 보여준다.

표 2. 데이터추출항목과 용도

추출항목	출처	용도	관련자료
제목, 저자명	해당논문	분류, 중복체크	–
출판연도	“	분류	그림4
문헌종류	문헌제공처	연구경향분석	그림15
주요연구주제	해당논문내용 (표2)	주요 연구 주제별 경향분석	그림4, 그림7, 그림8
세부연구주제	해당논문내용 (표2)	세부 연구 주제별 경향분석	그림5, 그림6, 그림9
문헌제공처의 인용횟수	문헌제공처	인용횟수에 따른 연구 경향분석	그림7, 그림8, 그림9
실험에 필요한 트윗 건수	해당논문내용	트윗건수에 따른 연구 경향분석	그림11
실험에 사용한 사용자 계정 수	해당논문내용	사용자 계정 수에 따른 연구 경향분석	그림10
실험에 필요한 트윗 샘플링 기간	해당논문내용	샘플링 기간에 따른 연구 경향분석	그림12
저자의 전공, 소속국가, 소속기관, 저자수	해당논문	저자의 특성 분석	그림13, 그림14

표 3. 주요연구주제와 세부연구주제

주요연구 주제	세부연구주제
응용 (Applications)	전자상거래(Commerce), 기업(Company) 위기 커뮤니케이션(Crisis communication) 클라우드센싱(Crowd-sourced sensing) 재해예방(Disaster prevention) 교육(Education), 사업(Entrepreneur) 헬스케어(Health care), 뉴스(News) 정치(Politics),공공관계(Public relations) 로보틱스(Robotics), 증권시장(Stock market)
효과적사용 (Effective Use)	정보신뢰성분석(Information credibility analysis) 기술경영(Technology Management) 트위터네트워크분석(Twitter network analysis)
사용자의 사용경향 (Tendency among users)	아동의사용(Adolescent' s use) 비공식통신수단(Informal communication) 트위터소개(Introduction) 리트윗(Retweet), 대화주제검색(Topic detection) 트위터사용경향(Trend), 언팔로우(Unfollow)
트윗내용분석 (Tweet Content Analysis)	사건탐지(Event detection) 정보필터링(Information filtering) 저널리스트분석(Journalist analysis) 위치정보탐색(Location information) 의견/감성분석(Opinion/Sentiment analysis) 사용자성격예측(Personality) 인기메시지(Popular message), 랭킹(Ranking) 추천(Recommendation), 사회적영향력(Social impact) 스팸탐지(Spam detection)
트윗텍스트 분석 (Tweet Text Analysis)	태크및@표시 분석(Tag/@sign analysis) 분류자분석(Classifier & labeler) 어휘표준화(Lexical normalization) 텍스트품사분석(Text Analysis)
사용자활동 (User activity)	정보생산자탐지,가상청중(Imagined audience) 정보전이(Information contagion) 정보흐름(Information flow) 대화흐름모델링(Modeling dialogue Action) 트위터사용(Twitter using) 사용자모델링(User modeling)
사용자관계성 (User relationships)	커뮤니티구조(Community structure) 팔로우관계(Follow relationship) 인플루언서(Influencer/User influence) 파워유저(Power user) 소셜네트워크,상호작용패턴(Social network) 관계형성(Tie formation)

본 연구의 가장 큰 목적은 트위터 관련 연구의 주요 연구 주제의 경향 파악이다. 본 연구에서 트위터 관련 연구의 주요 주제를 [표 3]과 같이 7 가지로 분류한다. 주요 연구 주제는 문헌의 조사의 결과로 파악할 수 있다. 우선, 조사 대상 연구문헌의 내용을 분석하여 주제에 따라 크게 분류한다. 이는 본연구의 주요 연구 주제가 된다. 다음으로 주요 연구 주제 내의 문헌을 다시 한번 내용에 따라 세부적으로 분류한다. 이는 본연구의 세부 연구 주제가 된다. “Application” 연구 분류는 트위터를 특정한 목적에 응용하고자 하는 목적을 가진 연구들의 분류이다. 정치[7][8], 경제[9][10], 교육[11], 건강[12] 등 다양한 분야의 응용에 관한 연구이다. “Effective Use”는 트위터의 효과적인 사용과 관련된 연구들의 분류로, 트윗 정보의 신뢰성 분석[13], 트위터 네트워크의 속성 분석 등에 관한 연구[14]가 있다. “Tendency among users”는 트위터 사용의 트렌드(trend)와 관련된 연구들의 분류로, 토픽 검색[15], 리트윗의 경향[16], 언팔로우의 경향[17], 트위터 사용의 경향성과 관련된 연구[18]가 있다. “Tweet Content Analysis”는 트윗의 내용을 분석하는 작업에 중점을 둔 연구 분류로, 사건 탐지[19], 위치정보 파악[20], 의견 또는 감성 분석[21-33], 스팸 탐지[23] 등에 관한 연구가 있다. “Twitter Text Analysis”는 트윗 문장의 문법적 분석을 실시하는 연구들의 분류로, 태그(tag)분석[24], 분류자(classifier)에 관한 연구[25] 등에 관한 연구가 있다. “User Activity”는 트위터 사용자의 활동에 중점을 둔 연구들의 분류로, 사용자 모델링[26], 정보 이동[27], 정보 흐름[28], 트위터 사용자의 활동성에 관한 연구[29]가 있다. “User Relationship”은 사용자의 관계성 분석에 중점을 둔 연구들의 분류로, 영향력있는 사용자(Influencer) 파악[30], 파워 유저[31], 관계 형성[32] 등에 관한 연구가 있다.

3. 문헌제공처 선정(Source Selection) 단계

문헌제공처 선정 단계는 관련 연구 문헌을 제공하는 문헌제공처를 선정하는 단계이다. 조사 대상 문헌은 Wikipedia[1]의 “List of academic database and search engines”의 리스트 중 자료의 풍부성, 접근용이성, 접근비용을 고려하여 선정한다. 선정된 문헌 제공처는 ACM Digital Library[34], Google Scholar[35], IEEE Xplore[36], Science Direct[37], Web of Science[38] 등 5개이다. 선택된 문헌 제공처에서 가장 좋은 검색결과를 제공하는 검색 키워드는 “Twitter”와 “tweet”이다. 문헌의 본문 등 자료의 모든 항목을 검색 대상으로 하면 너무 많은 결과가 산출되므로 키워드는 문헌의 제목에만 적용하도록 하였다. 문헌검색은 2012년 6월, 2013년 2월, 2014년 5월 등 3차례에 걸쳐 실시하였다. 검색

대상 문헌은 그룹 1과 그룹 2로 나누어 2007년부터 현재까지의 문헌을 그룹1으로 하고, 2011년부터 현재가지의 문헌을 그룹 2로 분류하였다. 각 문헌 제공처별 검색 결과는 [표 3]과 같다.

4. 조사대상 문헌 선정(Researches Selection) 단계

조사대상 문헌 선정 단계는 문헌제공처로부터 연구 경향을 분석하기 위한 조사대상 문헌을 선정하는 단계이다. 검색된 연구 문헌은 3단계에 걸쳐 대상을 압축하여 조사 대상 문헌을 선정한다. 조사 대상 선정 결과는 [표 4]와 같다. 단계별 선정 작업은 아래와 같다.

표 4. 문헌선정의 절차와 결과

문헌제공처	그룹	stage 1			stage 2	overlapping paper	final
		2012.6	2013.2	2014.5			
IEEE	그룹 1	113	196	404	11	4	8
	그룹 2	74	156	364	4	0	4
ACM	그룹 1	358	556	973	23	22	23
	그룹 2	223	358	824	14	11	14
GOOGLE	그룹 1	3,960	7,040	10,800	63	40	38
	그룹 2	1,680	3,210	6,750	21	16	13
SCIENCE	그룹 1	76	108	200	14	7	7
	그룹 2	44	76	167	4	4	1
WEB	그룹 1	262	368	702	25	18	14
	그룹 2	155	251	574	12	11	6
TOTAL	그룹 1 (2007~present)	4,769	8,268	13,079	136	91	90
	그룹 2 (2011~present)	2,176	4,051	8,679	55	42	38

(1) 1단계: 3차례에 걸쳐 문헌제공처의 검색엔진을 이용하여 키워드 “Twitter” 또는 “tweet”이 문헌의 제목에 나타나는 자료를 그룹 1과 그룹 2로 나누어 검색한다. 검색의 결과는 [표 4]의 “stage 1” 열에 나타나 있다.

(2) 2단계: 1단계의 검색결과, 각 문헌의 피인용수 등을 고려하여 각 문헌제공처의 대표가 될 만한 문헌을 선정한다. 검색의 결과는 [표 4]의 “stage 2”에 보여 진다. 문헌의 피인용수는 3회의 문헌검색마다 각 문헌의 피인용수를 파악하고 이를 평균한 수치를 그 문헌의 피인용수로 한다.

(3) 3단계: 다수의 문헌이 2개 이상의 문헌 제공처에서 검색된다. 중복 검색된 문헌의 수는 [표 4]의 “overlapping paper” 열에 보여 진다. 중복 검색된 문헌은 중요하다고 판단되어 가능한 한 조사 대상 문헌에 포함시킨다. 중복 검색 문헌의 수를 고려하여 최종적으로 조사 대상 문헌의 수를 결정한다. 결과는 [표 4]의 “final” 열에 보여진다. 따라서 그룹 1에서는 90개의 문헌, 그룹 2에서는 38개의 문헌 등 128개의 조사 대상 문헌이 선정된다. [그림 3]은 문헌 제공처별 선택된 논문의 중복을 보여준다.

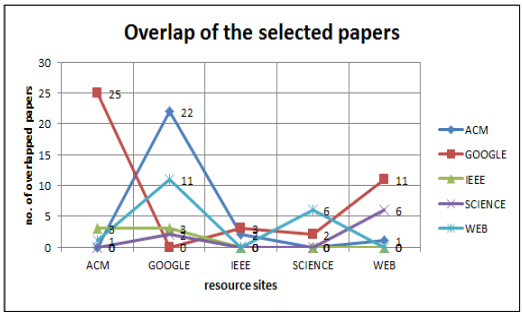

그림 3. 선택된 논문의 중복

5. 데이터 추출(Data Extraction) 단계

데이터 추출 단계는 조사 대상 문헌으로부터 연구 경향 파악을 위한 데이터를 추출하는 단계이다. 조사 대상 문헌을 면밀히 검토하여 각 문헌으로부터 3.2절에서 파악한 추출 대상 데이터 항목을 추출한다. [표 2]에서 표시된 출처로부터 해당항목을 추출하며, 추출한 데이터는 분석을 위해 컴퓨터 프로그램에 누적하고 다양한 자료로 가공한다.

Ⅳ. 문헌 조사 결과의 분석

[표 1]의 본 연구에서 제안한 SLR 기법에 의한 문헌 조사 결과가 분석된다.

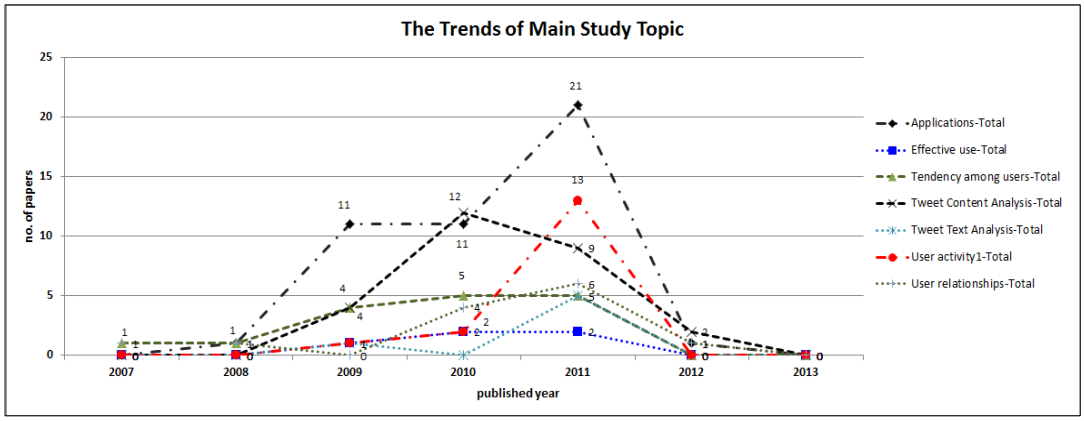

그림 4. 연도에 따른 주요 연구주제의 경향

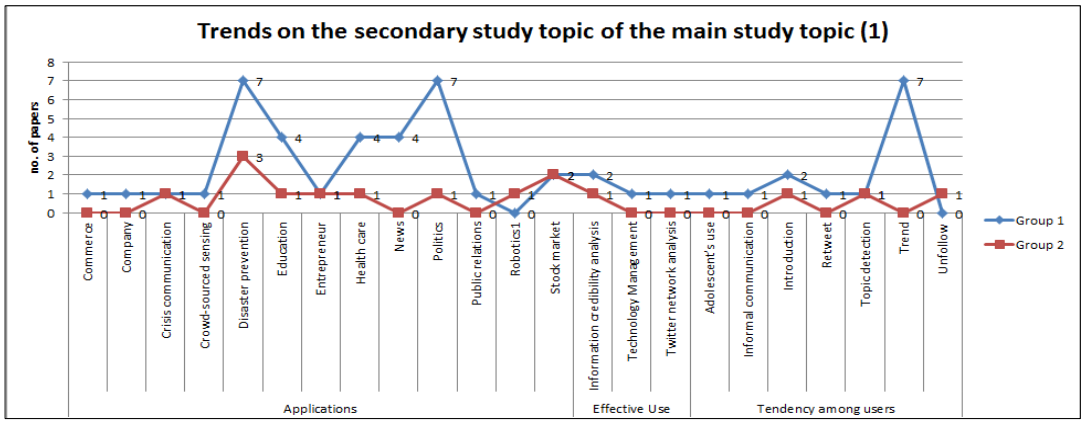

그림 5. 주요 연구주제별 세부 연구주제의 경향 (1)

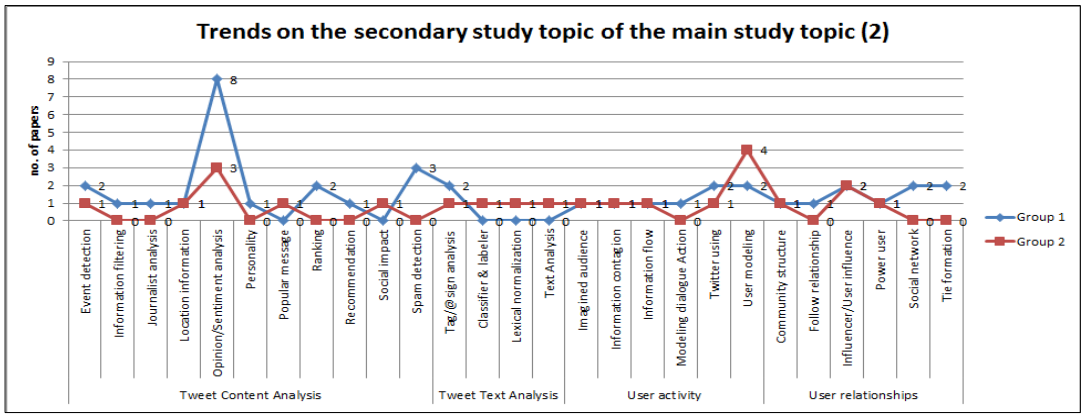

그림 6. 주요 연구주제별 세부 연구주제의 경향 (2)

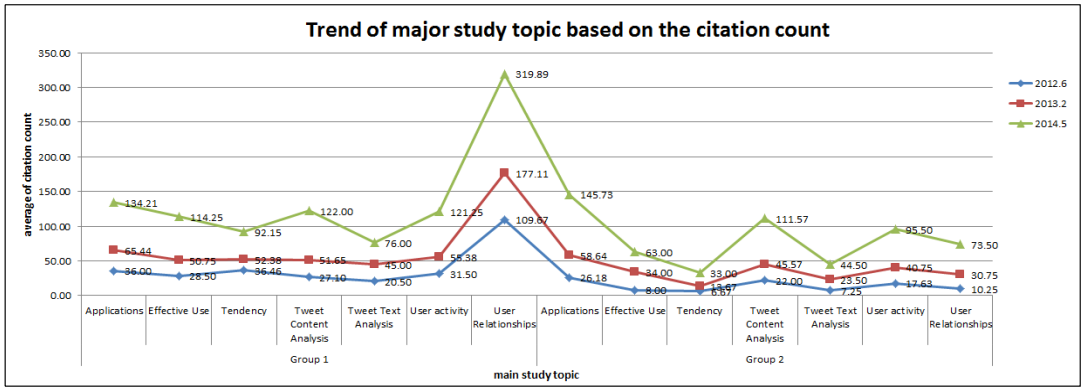

그림 7. 피인용수에 따른 주요 연구주제의 경향

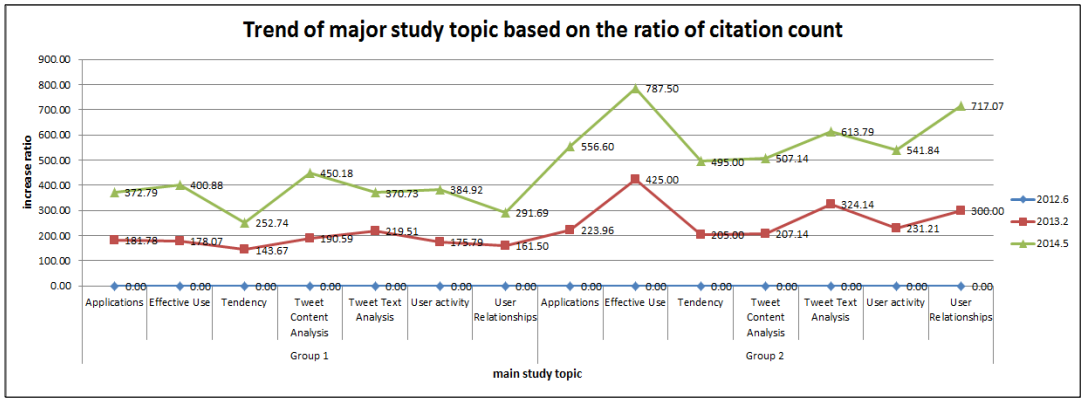

그림 8. 피인용수의 비율에 따른 주요 연구주제의 경향

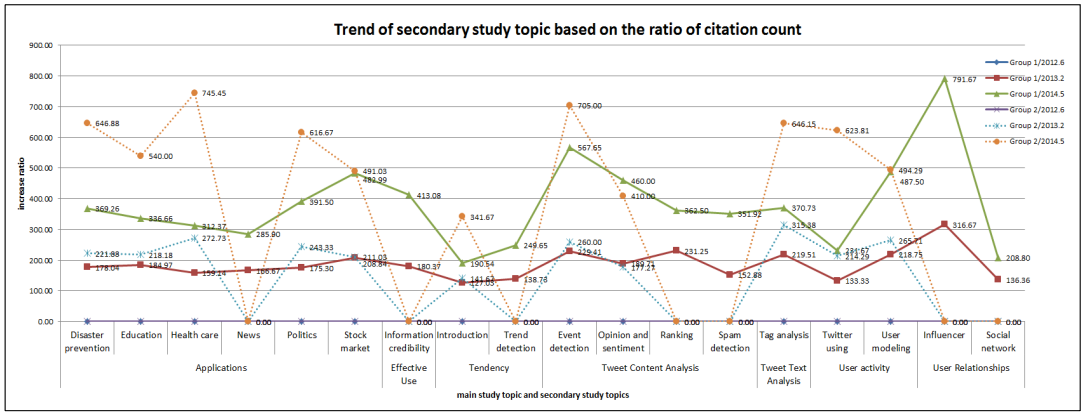

그림 9. 피인용수의 비율에 따른 세부 연구주제의 경향

1. 연도에 따른 주요 연구주제의 경향성

연도에 따른 주요 연구 주제의 경향성을 알아본다. 조사 대상 문헌의 연도별 주요 연구 주제에 관한 경향성은 [그림 4]에서 파악할 수 있다. 그림과 같이 다른 연구 주제와 비교하여 “Application”에 관한 연구가 상당히 많이 이루어지고 있음을 알 수 있고, “Tweet Content Analysis”에 관한 연구도 상당히 많이 이루어지고 있음을 알 수 있다. 또한 “User activity” 관련 연구도 비교적 많이 이루어지고 있음을 알 수 있다. 트위터 서비스가 시작된 이후 트위터 관련 주요 연구는 2009년에서 2011년 사이에 집중되어 많이 이루어졌음을 알 수 있다. 2012년과 2013년의 연구가 적은 이유는 조사 대상 문헌을 선정하는 과정에서 피인용수가 높은 연구를 중요한 연구로 선정하였기 때문인 것으로 분석된다.

주요 연구 주제별 세부 연구 주제에 관한 경향성을 살펴보면, [그림 5]와 같이 그룹 1의 경우, “Application”에는 “Disaster Prevention”, “Politics”, “Health Care” 그리고 “Education”과 관련된 연구가 많이 이루어지고 있음을 알 수 있고, “Effective Use”에는 “Information Credibility Analysis”와 관련된 연구가 많이 이루어지고 있음을 알 수 있다. “Tendency among users” 주요 연구 주제에 관해서는 “Trend”와 “Introduction”에 관한 세부 연구 주제가 많이 연구되고 있음을 알 수 있다. 그룹 2의 경우, “Application”의 “Disaster prevention”과 “Stock market”에 관해 비교적 많은 연구가 이루어지고 있음을 알 수 있다. [그림 6]을 살펴보면 그룹 1의 경우, “Tweet Content Analysis” 주요 연구 주제에서 “Opinion/Sentiment analysis”와 “Spam detection”에 관한 세부 연구가 많이 이루어지고 있음을 알 수 있고, 그룹 2의 경우, “Tweet Content Analysis” 주요 연구 주제 중 “Opinion/Sentiment analysis”가 많이 연구되고, “User activity”는 “User modeling”에 관한 연구가 비교적 많이 이루어지고 있음을 알 수 있다.

2. 피인용수에 따른 주요 연구주제의 경향성

피인용수에 따른 주요 연구 주제의 경향성을 알아본다. 조사 대상 문헌의 피인용수에 따른 주요 연구 주제에 관한 경향성은 [그림 7] 및 [그림 8]과 같다. [그림 7]은 주요 연구 주제별 피인용수의 증가를 보여주고 있다. 그룹 1의 “User Relationship”, “Application”, “Tweet Content Analysis” 그리고 “User Activity” 관련 연구의 피인용수가 높게 나타난다. 그룹 2에는 “Application” 과 “Tweet Content Analysis” 관련 연구의 피인용수가 높게 나타난다. 그러나 피인용수와 증가율의 경향은 일치하지 않는다. [그림 8]은 주요 연구 주제의 피인용수의 증가 비율을 보여주고 있다.

[그림 8]과 같이 그룹 1에서는 “Tweet Content Analysis”와 “Effective Use” 관련 연구의 피인용수가 비교적 많이 증가하고 있음을 보여주고 있다. 그룹 2에서는 “Effective Use”, “User Relationship” 그리고 “Tweet Text Analysis” 관련 연구의 피인용수가 빠르게 증가하고 있음을 보여준다.

피인용수에 따른 세부 연구 주제의 경향성을 알아보면, [그림 9]와 같이 “Application” 주요 연구 주제에서는 “Stock market” 과 “Education” 관련 연구의 피인용수가 비교적 많이 증가하고 있음을 보여주고 있다. “Tweet Content Analysis” 주요 연구 주제에서는 “Opinion and sentiment” 관련 연구의 피인용수가 비교적 많이 증가하고 있고, “User Relationships” 주요 연구 주제에서는 “Social network” 관련 세부 연구 주제의 피인용수가 비교적 많이 증가하고 있음을 알 수 있다. 피인용수에 따른 세부 연구 주제의 경향성은 그룹 1과 그룹 2가 유사하게 나타나고 있다.

3. 실험자료의 규모에 관한 분석

3장에서는 조사 대상 문헌 내에서 각 연구의 실험에서 사용한 트윗의 규모, 사용자 계정의 수, 트윗의 발생 기간을 조사하였다. 이와 같은 자료를 주요 연구 주제와 연관하여 주요 연구 주제의 특성을 파악하였다. [그림 10]은 각 주요 연구 주제별 실험에 사용한 트윗터 사용자 계정수를 보여주고 있다. 그림과 같이 그룹 1에서는 “Tweet Content Analysis”의 “Tag analysis”, “Spam detection” 그리고 “Opinion/Sentiment analysis” 연구와 “Application”의 “Disaster prevention” 관련 연구에서 비교적 많은 사용자 계정을 사용하는 것으로 나타

났다. 그룹 2에서는 “Application”의 “Disaster prevention”, “Tweet Content Analysis”의 “Opinion/Sentiment analysis” 그리고 “User Relationships”의 “Influencer” 관련 연구가 비교적 많은 사용자 계정을 사용하는 것으로 나타났다. [그림 11]은 각 주요 연구 주제별 실험에 사용한 트윗의 규모를 보여주고 있다. 각 주요 주제별 트윗의 규모는 [그림 10]의 사용자 계정 수의 경향과 거의 일치한다. 즉, 사용자 계정을 보다 많이 사용한 주요 연구 주제들이 더 많은 트윗을 실험에 이용한 것으로 나타났다. 다만, 그룹 2의 경우 “Tendency”의 “User modeling”이 사용자 계정수와 상관없이 비교적 많은 트윗을 사용한 것으로 나타났다. [그림 12]는 샘플링한 트윗이 발생된 기간을 주요 연구 주제별로 표시한 것이다. 트윗 발생 기간은 [그림 11]의 트윗의 규모의 경향성과 상당히 일치함을 알 수 있다. 다만, “Tweet Content Analysis”의 “Opinion/Sentiment analysis” 연구의 경우 트윗의 규모는 크지만, 트윗 발생기간은 길지 않음을 알 수 있다. 트윗의 규모, 사용자의 계정수, 트윗의 발생 기간은 주요 연구 주제별 문헌들의 수치를 평균하여 구하였다. 하지만 규모를 밝히지 않거나 실험을 실시하지 않은 연구는 제외하였고, 수치를 알 수 있는 연구들의 평균만 구하였다.

4. 저자의 특성에 관한 분석

트위터 관련 연구의 주요 연구 주제와 아울러 각 연구의 연구자의 특성을 파악해 본다. [그림 13]은 조사 대상 문헌의 주저자의 전공을 파악한 결과이다. 그림과 같이 저자들의 전공은 컴퓨터과학 (Computer science)이 가장 많았고, 정보 (information), 저널리즘 (journalism) 그리고 Yahoo, Microsoft와 같은 연구기관(research)의 연구자도 비교적 다수였다. 저자의 전공은 제1저자의 소속부서를 기준으로 파악하였다. [그림 14]는 저자의 소속 국가를 보여주고 있다. 그림과 같이 조사 대상 문헌 저자의 소속 국가는 미국이 가장 많았다. [그림 15]는 조사 대상 문헌의 종류를 파악한 것이다. 프로시딩(proceeding) 형태의 논문과 논문지 (journal)에 게재된 논문이 다수였다.

5. 연구의 경향성 요약

위와 같은 문헌 조사 결과의 분석으로 다음과 같은 몇 가지 트위터 관련 연구의 경향성을 파악할 수 있다.

첫째, 연도별 주요 연구 주제의 경향성과 관련하여, 2007년 이후 트위터 관련 연구는 “Application”, “Tweet Content Analysis” 그리고 “User activity” 주제와 관련한 연구가 가장 많이 이루어지고 있고, 주요 연구는 2009년에서 2011년 사이에 많이 이루어지고 있음을 알 수 있다.

둘째, 주요 연구 주제별 세부 연구 주제의 경향성과 관련하여, “Application”에서는 “Disaster Prevention”, “Politics”, “Health Care” 그리고 “Education”과 관련된 세부 연구 주제가 많이 연구되고 있음을 알 수 있다. “Tweet Content Analysis”에서는 “Opinion/Sentiment Analysis” 관련 연구가 많이 이루어지고, “Tendency among users” 주요 연구 주제에 관해서는 “Trend”와 관련한 주제가 많이 연구되고 있음을 알 수 있다.

셋째, 피인용수에 따른 주요 연구 주제의 경향성과 관련하여, 그룹 1의 “User Relationship”, “Application”, “Tweet Content Analysis” 그리고 “User Activity” 관련 연구의 피인용수가 높음을 알 수 있고, 그룹 2에는 “Application” 과 “Tweet Content Analysis” 관련 연구의 피인용수가 비교적 높음을 알 수 있다.

넷째, 피인용수의 증가율에 따른 주요 연구 주제의 경향성과 관련하여, 그룹 1에서는 “Tweet Content Analysis”와 “Effective Use” 관련 연구의 피인용수가 비교적 많이 증가하고 있고, 그룹 2에서는 “Effective Use”, “User Relationship” 그리고 “Tweet Text Analysis” 관련 연구의 피인용수가 빠르게 증가하고 있음을 알 수 있다.

다섯째, 주요 연구 주제별 실험 자료의 규모와 관련하여, “Tweet Content Analysis”의 “Tag analysis”, “Spam detection” 그리고 “Opinion/Sentiment analysis” 연구가 비교적 많은 사용자 계정을 사용함을 알 수 있다. 또한 “Application”의 “Disaster prevention”, “User Relationships”의 “Influencer” 관련 연구도 비교적 많은 사용자 계정을 필요로 함을 알 수 있다. 주요 연구 주제별 트윗의 규모는 사용자 계정수

의 경향과 거의 일치하고, 트윗 발생 기간도 트윗의 규모의 경향성과 상당히 일치함을 알 수 있다.

여섯째, 저자의 특성과 관련하여, 트위터 관련 연구자의 전공은 컴퓨터과학 (Computer science)이 가장 많음을 알 수 있고, 정보 (information), 저널리즘 (journalism)관련 전공도 비교적 다수임을 알 수 있다. 저자의 소속 국가는 미국이 가장 많음을 알 수 있다.

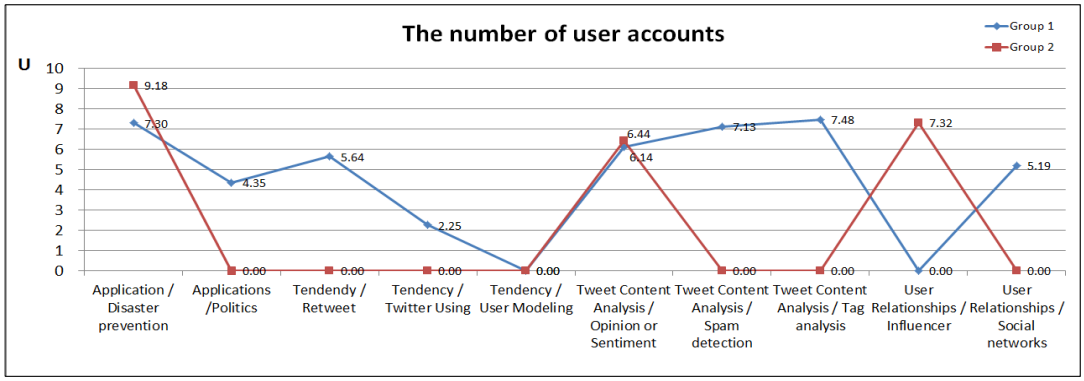

그림 10. 세부 연구주제의 사용자 계정수

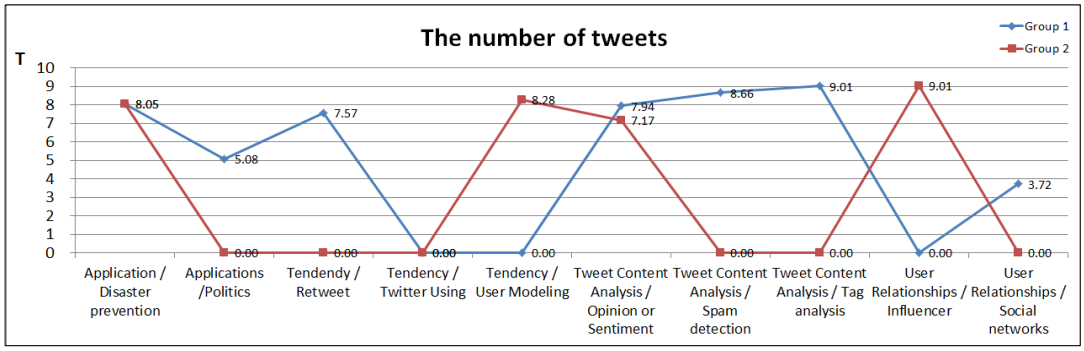

그림 11. 세부 연구주제의 트윗 수

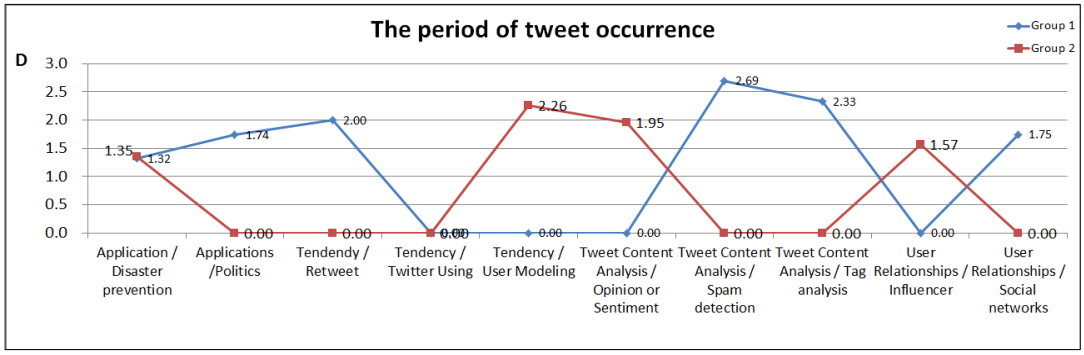

그림 12. 세부 연구주제의 트윗 발생 기간

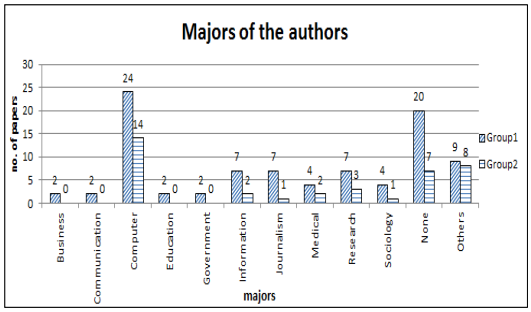

그림 13. 저자의 전공

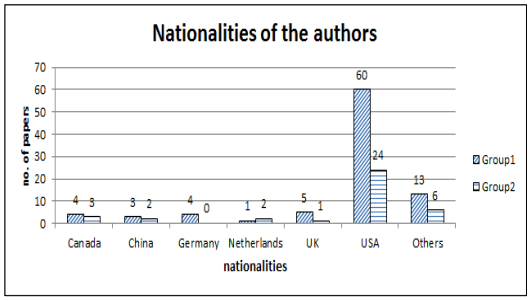

그림 14. 저자의 소속국가

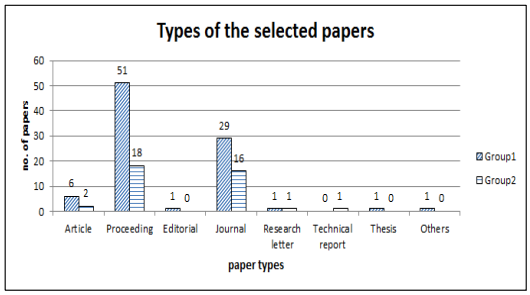

그림 15. 선택된 논문의 형태

V. 결 론

소셜 네트워크 서비스는 다양한 이점 때문에 일상생활에서 편리한 의사소통 도구로 사용되고 있으며 이를 다양한 분야에 활용하고자 하는 연구도 활발히 진행되고 있다. 연구의 진행과 함께 관련 연구를 조사하고 경향성을 파악하는 작업은 필수적이다. 하지만 방대한 양의 문헌을 조사하는 것은 여러 가지 측면으로 어려운 작업이다. 따라서 본 연구에서는 소셜 네트워크 서비스 중 트위터관련 연구에 중점을 두어 방대한 양의 트위터 관련 문헌을 조사하여 연구의 경향성을 밝힌다. 특히 체계적인 문헌 조사를 위하여 SLR 기법을 도입하여 6단계에 걸친 체계적 문헌 조사가 이루어지도록 한다. 문헌 조사를 위해 보편적으로 잘 알려진 5개 문헌 제공처로부터 문헌을 검색하여 최종적으로 128개의 대표적인 문헌을 선정하였다. 조사 대상 문헌으로부터 연구의 의문점을 해결할 수 있는 자료를 추출하고 이를 취합하여 조사 분석 작업을 실시하였다. 조사 분석 결과로 4.4절에 요약한 것과 같은 트위터 관련 연구와 관련한 몇 가지 중요한 연구의 경향성을 얻을 수 있었다.

참 고 문 헌

[1] WIKIPEDIA: http://www.wikipedia.org/ (accessed May 2, 2014)

[2] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," in Proceeding of the 19th international conference on World wide web, pp.591-600, 2010.

[3] B. Kitchenham, "Procedures for Undertaking Systematic Reviews," Joint Technical Report, Computer Science Department, Keele University (TR/SE-0401) and National ICT Australia Ltd. (0400011T.1), 2004.

[4] F. Lucas, F. Molina, and A. Toval, "A systematic review of UML model consistency management," Information and Software Technology, Vol.51, pp.1631-1645, 2009.

[5] D. Budgen, A. J. Burn, O. P. Brereton, B. A. Kitchenham, and R. Pretorius, "Empirical evidence about the UML: a systematic literature review," Software: Practice and Experience, Vol.41, Iss.4, pp.363-392, 2011.

[6] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage

and communities," in Proceedings of the 9th WebKDD, pp.56-65, 2007.

[7] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, pp.178-185, 2010.

[8] M. Conover, J. Ratkiewicz, M. Francisco, B. Goncalves, A. Flammini, and F. Menczer, "Political polarization on twitter," in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp.89-96, 2011.

[9] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," Journal of Computational Science, Vol.2, Iss.1, pp.1-8, 2011.

[10] J. Yoon and B. Hwang, "Design of a Productivity Measuring System for Twitter Marketing Reflecting User"s Sentiment in Social Networks," Journal of KISS: Database, Vol.40, No.6, pp.377-385, 2013.

[11] R. Junco, G. Heiberger, and E. Loken, "The effect of Twitter on college student engagement and grades," Journal of Computer Assited Learning, Vol.27, Iss.2, pp.119-132, 2011.

[12] C. Hawn, "Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care," Health Affairs, Vol.28, No.2, pp.361-368, 2009.

[13] C. Castillo, M. Mendoza, B. Poblete, "Information credibility on twitter," in Proceedings of the 20th international conference on World wide web, pp.675-684, 2011.

[14] A. Teutle, "Twitter: Network properties analysis," in Proceeding of 20th International Conference on Electronics, Communications and Computer(CONIELECOMP), pp.180-186, 2010.

[15] M. Cataldi, L. Caro, and C. Schifanella, "Emerging topic detection on twitter based on temporal and social terms evaluation," in Proceedings of the Tenth International Workshop on Multimedia Data Mining, 2010.

[16] D. Boyd, S. Golder, and G. Lotan, "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter," in Proceeding of 43rd Hawaii International Conference on System Sciences (HICSS), pp.1-10, 2010.

[17] H. Kwak, H. Chun, and S. Moon, "Fragile online relationship: a first look at unfollow dynamics in twitter," in Proceeding of the SIGCHI Conference on Human Factors in Computing Systems, pp.1091-1100, 2011.

[18] M. Mathioudaks, and N. Koudas, "TwitterMonitor: trend detection over the twitter stream," in Proceeding of the 2010 ACM SIGMOD International Conference on Management of data, pp.1155-1158, 2010.

[19] H. Becker, M. Naaman, and L. Gravano, "Beyond trending topics: Real-world event identification on Twitter," in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp.438-441, 2011.

[20] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," in Proceedings of the 19th ACM international conference on Information and knowledge management, pp.759-768, 2010.

[21] A. Pak, and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in Proceeding of 7th international conference on Language Resources and Evaluation (LREC 2010), pp.1320-1326, 2010.

[22] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment

and socio-economic phenomena," in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp.450-453, 2011.

[23] S. Yardi, D. Romero, G. Schoenebeck, and D. boyd, "Detecting spam in a twitter network," First Monday, Vol.15, No.1-4, 2010.

[24] D. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter," in Proceedings of the 20th international conference on World wide web, pp.695-704, 2011.

[25] X. Liu, S. Zhang, F. Wei, and M. Zhou, "Recognizing named entities in tweets," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp.359-367, 2011.

[26] M. Perc, "Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number," PLoS ONE, Vol.6, No.8, e22656, 2011.

[27] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on Digg and Twitter social networks," in Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, pp.90-97, 2010.

[28] G. Lotan, E. Graeff, M. Ananny, D. Gaffney, I. Pearce, and D. Boyd, "The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions," International Journal of Communication, Vol.5, pp.1375-1405, 2011.

[29] W. Reinhardt, M. Ebner, G. Beham, C. Costa, "How people are using Twitter during conferences," in Proceeding of 5. EduMedia conference, pp.145-156, 2009.

[30] E. Bakshy, J. Hofman, W. Mason, and D. Watts, "Everyone's an influencer: quantifying influence on twitter," in Proceedings of the fourth ACM international conference on Web search and data mining, pp.65-74, 2011.

[31] S. Wu, J. Hofman, W. Mason, and D. Watts, "Who says what to whom on twitter," in Proceedings of the 20th international conference on World wide web, pp.705-714, 2011.

[32] Y. Takhteyev, A. Gruzd, and B. Wellman, "Geography of Twitter networks," Social Networks, Vol.34, Iss.1, pp.73-81, 2012.

[33] G. Lee and K. Lee, "Twitter Sentiment Analysis for the Recent Trend Extracted from the Newspaper Article," KIPS transactions o software and data engineering, Vol.2, No.10, pp.731-738, 2013

[34] IEEE Xplore: http://ieeexplore.ieee.org/ (accessed May 2, 2014)

[35] ACM Digital Library:http://dl.acm.org/(accessed May 2, 2014)

[36] Google Scholar: http://scholar.google.co.kr/(accessed May 2, 2014)

[37] Science Direct: http://www.sciencedirect.com/(accessed May 2, 2014)

[38] Web of Science: http://apps.webofknowledge.com/ (accessed May 2, 2014)

[39] O. Phelan, K. McCarthy, and B. Smyth, "Using Twitter to Recommend Real-Time Topical News," in Proceedings of the third ACM conference on Recommender systems, pp.385-388, 2009

[40] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," in Proceedings of the 19th international conference on world wide web, pp.851-860, 2010

저 자 소 개

하 일 규(IlKyu Ha) 정회원

- 1992년 2월 : 영남대학교 전산공학과(공학사)
- 2001년 2월 : 영남대학교 정보처리교육전공(교육학석사)
- 2003년 8월 : 영남대학교 컴퓨터공학과(공학박사)
- 1992년 1월 ~ 1995년 10월 : 증권감독원 전산실
- 2002년 3월 ~ 현재 : 영남대학교 강사, 객원교수

<관심분야> : 소셜 네트워크 분석, 빅데이터 처리 및 분석, 인체 센서 네트워크