
빅데이터를 통해 본 한국 현대 시인 인지도 및 평판분석*

최도식 (강원대학교)

I. 머리말

“한국을 대표하는 시인은 누구일까?” 우리는 가끔씩 이 물음에 대한 답변을 궁금해 한다. 더욱이 한국을 대표할 수 있는 한 명의 시인을 꼽으라면 누구를 거명(擧名)해야 할까? 1920년대 자유시가 본격적으로 창작된 이후 현대시는 거의 100년이란 세월의 족적을 남겼다. 그 현대 시사의 족적을 더듬어 보면 많은 현대 시인들이 있었고 그 시인들에 의해 새로운 풍과 기법의 시들이 창작되었다. 그런데 대중에게 자주 이름이 오르내리는 시인들이 있는가 하면 잊어진 시인들도 있다. 이런 시점에서 우리는 적어도 한국 현대 시인들 중 누가 대중에게 자주 이름이 오르내리는 시인인가를 되돌아보아야 할 때이다. 특히, 한국의 대중이 자주 거명하는 시인이 누구이며, 그 시인의 어떤 작품이 대중의 사랑을 받고 있는가를 살펴볼 필요가 있다. 그런데 이 문제를 규명하기란 쉬운 일이 아니다. 하지만 최근 빅데이터의 등장은 이 문제를 규명하는데 한걸음 나아갈 수 있는 진전을 보였다.

지금 전 세계를 뒤흔들고 있는 4차 산업 혁명, 빅데이터, 인공지능, 사물인터넷 등은 데이터에 기반을 두고 있다. 이 새로운 기술 혁명은 2016년 3월 알파고와 이세돌 간 세기의 바둑 대결에서 이세돌의 패배로 인공지능에 대한 관심이 급부상하면서 세간에 주목을 받았다. 대중에게 잘 알려져

1. 2016년도 강원대학교 대학회계 학술연구조성비로 연구하였음.(관리번호-620160141)

있지 않던 빅데이터와 인공지능은 4차 산업 혁명을 주도할 새로운 기술로 각광을 받았다.² 그러나 학계에서는 이전부터 빅데이터에 주목해왔다. 그리고 최근에 빅데이터와 인문학은 융합의 길을 모색하고 있다. 특히, 2016년 2월 『시민인문학』은 “빅데이터, 인문학 미래를 위한 디딤돌”이라는 주제로 빅데이터와 인문학의 응전(김기봉, 2016), 역할과 과제(김연권, 2016), 융합(김호연, 2016), 인문학적 상상력(유강하, 2016), 문학적 활용과 성찰 등을 특집 논문으로 다룬다. 이들 논문 중 정과리의 「빅데이터의 문학적 활용과 빅데이터에 대한 문학적 성찰」은 한국인의 ‘애(哀)’에 관한 감정이 우울이 아니라 설움이었음을 밝혀낸다. 그런데 그 방법론은 1990년 이후부터 2014년까지 출간된 시집 중 무작위적으로 298개의 시집에서 ‘우울’을 포함한 시집과 어휘의 빈도수를 추출하였으며, 해방 전 300권의 시집에서 ‘우울’이란 단어를 추출한 방식이었다. 그런데 이 방법론은 아날로그적 방식에 따른 데이터 추출의 결과이다. 이로 인해 정과리는 문학 텍스트를 빅데이터로 활용할 수 있는 디지털화의 공정에 대한 필요성을 제기한다. 즉, 문학 텍스트마이닝 공정에 대한 필요성을 제기한다(정과리, 2016, 117-133).

현재 문학 텍스트마이닝은 서지계량학적 연구에서 주로 사용되고 있는 토픽모델링분석(박자연·송민, 2013)과 저자동시인용분석이 있으며(김희전·조현양, 2010), 개체계량학적 분석(박진균·김택윤·송민, 2017) 등이 있다. 그리고 한국 현대시 코퍼스(KoPoCo) 구축으로 시어의 계량적, 통계적 빈도 분석이 이루어지고 있다(김병선, 2006).³ 그런데 이 연구들은 한정된 시인 연구 혹은 시어 연구이다. 곧 대중의 인지에 따른 시인이나 작품 검색, 시인 평판작품 평판이 아니다. 따라서 본 연구는 일반 대중이 바라본

-
2. 『MGI』 보고서에서는 ‘빅데이터’는 기존의 일반적인 데이터베이스 소프트웨어 틀로는 캡처, 저장, 관리 및 분석 할 수 없는 데이터셋을 넘어서는 규모의 데이터를 의미하며, 혁신경쟁 생산성을 위한 차세대 프론티어로 정의하고 있다(McKinsey, 2011).
 3. 김병선은 현대시 데이터베이스 구축을 주요 연구로 삼아 현대시의 계량적 문체 연구(김병선, 2000, 10-27), 현대시어 용례사전 등 현대시 시어 데이터베이스를 구축하고 있다.

한국 현대 시인에 대한 인지(도)와 인식(평판) 분석을 연구 목적으로 삼고자 한다. 그리고 그에 따른 결과를 통해 대중이 바라는 한국 현대 시인의 초상(肖像)을 예측해 보고자 한다.

II. 데이터 수집 방법과 대상 선정

1. 데이터 수집과 분석 방법

본 연구는 빅데이터를 기반으로 한 구글 검색량 분석과 썬트렌드, 소셜메트릭스를 이용한 오피니언 마이닝을 통해 데이터를 수집하였다. 먼저 구글 검색량 분석은 구글 트렌드(trends.google.com)를 활용하였다.⁴ 구글은 구글 트렌드를 통해 2004년 1월부터 지금까지의 검색어에 대한 검색량을 제공하고 있다. 실제 서비스는 2008년부터 시작되었으며, 국가별, 지역별, 언어별 등 특정 검색어에 대한 검색 빈도를 그래프로 제공하고 있다. 또한 구글 트렌드는 건강·게임·과학·금융·뉴스·도서 및 문학 등 다양한 카테고리에 따른 정보를 분리·통합하여 제공하고 있으며, 웹 검색을 기반으로 한 이미지 검색, 뉴스 검색, 유튜브 검색 등 특정 검색을 분리 검색할 수 있다. 따라서 본 연구는 2008년부터 현재까지의 검색량을 검색 기간으로 설정하였다. 그리고 관련 주제와 관련 검색어를 시인별로 우선 분석하였다. 설정 기간 내 대상 시인별 검색량을 1차 분석 대상으로 삼았으며, 검색량이 절대치 25를 일정 정도 선회하는 27인의 시인을 대상으로 2차 비교 분석하였다. 2차 비교 분석은 김소월을 기준 검색어로 삼아 타 시인과의 비교 분석을 실행했다. 김소월을 기준 검색어로 한용운, 서정주, 김수영 등 일정 검색 빈도가 추출된 시인을 대상으로 작고 시인 5인과 기성 시인 5인을 다시 2차 비교 분석하였다.

4. 『구글 트렌드』, 구글, <https://trends.google.com>

평판분석은 구글 검색량 비교 분석의 대상인 15인의 시인을 대상으로 분석했다. 분석 도구로는 다음소프트에서 제공하는 썬트렌드(sometrend (some.co.kr/bigdata))⁵와 소셜메트릭스⁶를 이용해 연관어 맵을 활용한 감성분석, 언급량에 따른 긍정·중립·부정 반응 및 추이를 각각 분석했다. 그리고 15인의 대상 시인에 대한 감성분석의 결과에 따른 연관어 빈도와 언급량에서 긍정 반응이 많은 시인과 부정 반응이 많은 시인의 연관어를 비교 분석했다. 그리고 분석 결과에서 이상 현상으로 감지되는 반응을 추가적으로 분석하였다. 그리고 그에 따른 결과로 대중에게 비쳐진 한국 현대 시인의 상(像)을 예측했다.

2. 대상 시인 선정

한국은 아직 노벨문학상 수상자를 배출하지 못했다. 그런데 근래 고은 시인이 노벨문학상 후보로 거론되기도 하였지만 그것은 한국 내 분위기일 뿐이었다. 그리고 최근 문덕수 시인의 시집 『우체부』가 스웨덴어로 번역 출간되었고, 이것이 스웨덴 우메오의 ‘테그스 비아고드’에서 선보이자(조선일보, 2014.7.1) 문덕수 시인을 노벨문학상 후보 한국 시인으로 거론하는 문학지들도 있었다.⁷ 이처럼 한국의 노벨문학상 수상에 대한 열망은 유독 요란하다. “한국 대중은 노벨문학상 수상에 대한 기대를 갖고 있다. 하지만 한국문학의 성과에도 불구하고 여전히 기대에만 그치는 이유가 있다.”(2014.7.1) 그 주된 이유로 대개 한국문화와 유럽문화의 문화적 이질성에 따른 문학의 이해 불충분, 한국문학의 해외 번역 문제 등을 거론한다.

5. 『썬트렌드』, 다음소프트, <http://www.some.co.kr/>

6. 『소셜 메트릭스』, 다음소프트, <http://www.socialmetrics.co.kr/socialSearch.html>

7. 시 전문계간지 『한국시학』은 2012년 ‘한국시학’이 선정한 노벨문학상 후보 시인으로 문덕수 시인을 거명했다.(『한국시학』(2012.6, 한국시학사) 그리고 『한국문학시대』는 2017년 봄호에 한국문단의 원로 문덕수 시인의 장시집 『우체부』를 앞세워 「한국 노벨문학상, 때가됐다」 특집을 다룬다.(김용재(2017.3), 「문덕수 시인 『우체부』의 국제적 위상」, 『한국문학시대』, 대전문인총연합회)

그런데 이보다 심각한 문제는 한국 대중이 외국인으로부터 한국의 대표적 문학 작품이나 대표 시인이 누구인가라는 질문을 받았을 때 그들에게 해 줄 마땅한 답변이 없다는 것이다. “노벨문학상을 수상할 만한 한국 작가가 있는가?”라는 질문에는 ‘모른다’가 62%로 대다수를 차지했다. ‘없다’도 34%로 높게 나왔으며 ‘있다’는 응답은 4%에 불과했다. 문학 수준은 높지만 정작 노벨상을 탈 만한 사람은 모르거나 없다는 것이다. 수상할 만한 작가로는 고은·박완서·김소월·윤동주·김유정·염상섭 등이 거론됐다. 고은 시인을 제외하고는 거의가 작고한 문학인이다.”(정유진·손민지, 2016.10.31) 『한국인만 모르는 다른 대한민국』의 저자 임마누엘 페스트라이쉬는 “한국인 자신도 한국문학에 관심이 없는 경우가 많다. 예를 들어 외국인에게 한국의 대표적 문학 작품으로 춘향전을 소개하면서도 정작 본인은 춘향전을 완독하지 않았고 그리 깊은 관심을 두지 않는 사람이 많다. 즉 한국 문화와 문학을 소개하는 일에만 열을 올릴 뿐 한국 문화와 문학 그 자체에는 관심이 없”다고 지적한다(2013). 그의 언급에 따르면 한국 대중은 한국문학에 관심이 없다. 한국의 대표적 문학 작품이 어떤 작품이며, 대표적 작품으로 꼽을 수 있는 진정한 이유가 무엇인지에도 관심이 없다. 또한 한국의 대표적 시인으로 누구를 꼽을 것인가에도 관심이 없으며, 본인이 알고 있는 시인마저도 대학입시 준비로 중고교 때 배운 시인이 전부이다. 그런데 그마저도 시간이 지나 기억에 남은 시인이 몇 안 된다. 이러한 실정에서 노벨문학상 수상을 바란다는 것은 열망만 있을 뿐 현실을 도외시한 결과일 뿐이다. 따라서 본고는 구글 트렌드 검색량을 통해 대중이 많이 검색한 시인에는 어떤 시인이 있으며, 그 시인들의 작품에는 어떤 작품이 거론되고 있는가를 밝히고자 한다. 이를 위해 본 연구는 대상 시인들을 우선적으로 선정했다.

분석 대상은 첫째 독자의 시집 선택 기준(독자인지성), 둘째 일간지와 문예지의 대표 신인 선정 자료(대표성), 셋째 『한국 전후 문체시인 연구』의 연구 대상 시인(연구성)을 우선 대상 선정 시인으로 선정했다. 첫째, 분석 대상 시인(독자인지성)은 시집 선택 기준이 갖는 특성 중 시인의 인지도에 의해 시집을 구입한다는 설문에 따라 대상 시인을 우선적으로 선정하였다.

특히 10대의 경우 다른 연령대에 비해 시인의 인지도에 의해 시집을 구입한다는 응답이 많았으며, 이는 전 연령대에 걸쳐 시인의 인지도에 의해 시집을 구입한다는 응답이 높게 나타나고 있다. 따라서 본 연구의 대상 시인을 우선적으로 이상, 김소월, 천상병, 윤동주, 한용운, 박목월, 김춘수, 백석, 이육사, 신동엽, 안도현, 황지우, 이해인 등 중고교 교과서와 참고서 등에 등장하는 시인을 우선 대상으로 선정하였다.(임현순, 2008, 320-321) 그리고 둘째, 대표성은 1998년 『조선일보』가 ‘대한민국 50년’이라는 일련의 기획 중 「문학 50년」이라는 제목 아래 ‘건국 이후 뛰어난 시인 49인’을 발표한 49인의 시인을 분석 대상 시인에 포함시켰다(1998.7.31). 여기에 더해 문학세계사에서 발행하는 시 전문 계간지 『시인세계』 창간호에서 100명의 현역 시인과 평론가들이 뽑은 「현대시 100년 10명의 시인」이라는 특집에 실린 10인의 시인을 분석 대상 시인에 포함했다.(권혁용, 전운호, 2002, 27-33) 그리고 셋째, 연구성은 「1970년 한국 현대시의 계보와 유형 연구를 위한 시론」에서 밝힌 리얼리즘시의 시적 계보, 모더니즘의 시적 계보, 전통서정시의 시적 계보에 해당하는 시인 중 분석 대상에서 배제된 시인을 분석 대상 시인에 포함시켰다.(류찬열, 2008, 203-225) 마지막으로 『한국 전후 문체시인 연구』(김학동, 2005)에서 연구된 정지용, 김기림, 김영랑, 김광섭, 신석초, 서정주, 유치환, 조지훈, 박두진, 오장환, 모운숙, 노천명, 정한모, 이형기를 추가적으로 포함시켰으며, 전후 시인으로 구상, 한하운, 김현승, 김종삼, 박인환, 김수영, 박봉우, 송옥, 조병화 등을 포함시켜 분석 대상 시인으로 삼았다. 그런데 여기에서도 배제된 기성 시인들이 있어 도중환, 김용택, 서정운, 박남준, 장정일, 문덕수, 강은교, 유안진 등을 추가적으로 분석 대상 시인에 포함시켰으며, 중복된 시인을 정리하여 총 100인의 시인을 개별적 분석 대상으로 삼아 분석을 실행했다.

본 연구의 개별적 분석 실행 대상 시인은 다음과 같다.

강은교, 고정희, 고은, 광재구, 구상, 기형도, 김경린, 김광규, 김광섭, 김규동, 김기린, 김기림, 김기택, 김남조, 김남주, 김동환, 김명인, 김소월, 김수영, 김승희, 김영랑, 김용택, 김종삼, 김지하, 김춘수, 김현승, 김혜순,

나태주, 나희덕, 노천명, 도종환, 류시화, 마종기, 모운숙, 문덕수, 박남수, 박남준, 박노해, 박두진, 박목월, 박봉우, 박세영, 박세현, 박용래, 박인환, 박재삼, 백무산, 백석, 서정윤, 서정주, 송수권, 송옥, 신경림, 신동엽, 신석정, 신석초, 안도현, 오규원, 오세영, 오장환, 유안진, 유진오, 유치환, 윤동주, 이가림, 이근배, 이동주, 이상, 이상화, 이성교, 이성복, 이성부, 이수익, 이승훈, 이시영, 이용악, 이장희, 이해인, 이형기, 임화, 장정일, 전봉건, 정지용, 정진규, 정한모, 정현중, 정호승, 정희성, 조병화, 조정권, 조지훈, 천상병, 최두석, 최승자, 최승호, 하종오, 한용운, 한하운, 황동규, 황지우(총 100인)

III. 현대 시인 구글 트렌드 검색량 비교 분석

본고는 총 100인의 시인을 대상으로 구글 트렌드를 활용해 각 시인들을 탐색했다. 탐색의 조건으로 검색어와 시인 주제어로 각각을 살폈으며, 검색어로 탐색했을 경우와 시인 주제어로 탐색했을 경우 두 탐색에서 다른 검색량을 보일 경우 시인 주제어를 분석의 결과로 가져왔다.

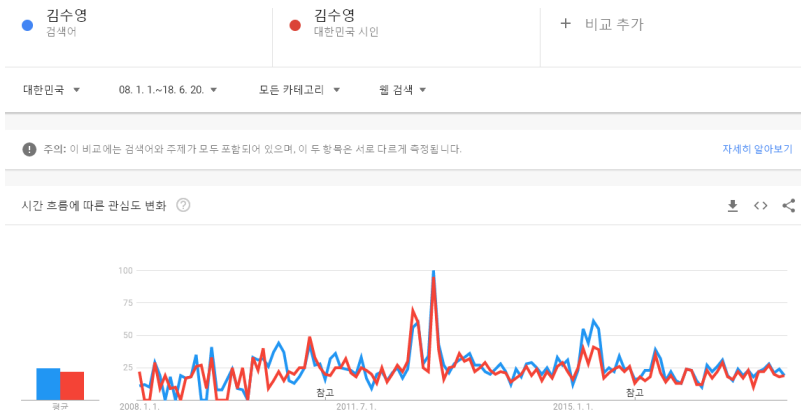


그림 1 김수영 탐색 조건

예를 들어, 그림 1은 김수영에 대한 탐색이다. 그림 1에서 파란색 그래프의 경우 ‘김수영 검색어’로 탐색한 조건이며, 붉은색 그래프는 ‘김수영 대한민국 시인’으로 탐색한 조건이다. 그런데 김수영의 경우 현격한 차이는 아니지만 평균값과 일정 부분에서 차이를 보이고 있다. 반면, 그림 2는 김소월에 대한 탐색이다. 마찬가지로 그림 2에서 파란색 그래프의 경우 ‘김소월 검색어’로 탐색한 조건이며, 붉은색 그래프는 ‘김소월 시인’으로 탐색한 조건이다. 그런데 김소월의 경우 평균값이 거의 차이가 없으며, 두 그래프도 동선을 따라 거의 비슷한 선을 이루며 변화하고 있다. 이와 같이 ‘검색어’ 탐색과 ‘시인’ 탐색으로 탐색 조건을 달리했을 때 두 탐색 조건이 일치하는 경우와 불일치하는 경우가 있다. 김수영의 경우 현격한 차이는 아니지만 평균값과 일정 부분에서 차이를 보인다. ‘김수영’에 대한 차이는 ‘검색어’ 조건에 개그맨 김수영과 사나리오 작가 김수영에 대한 관련 검색어가 포함됨으로써 나타난 차이이다.

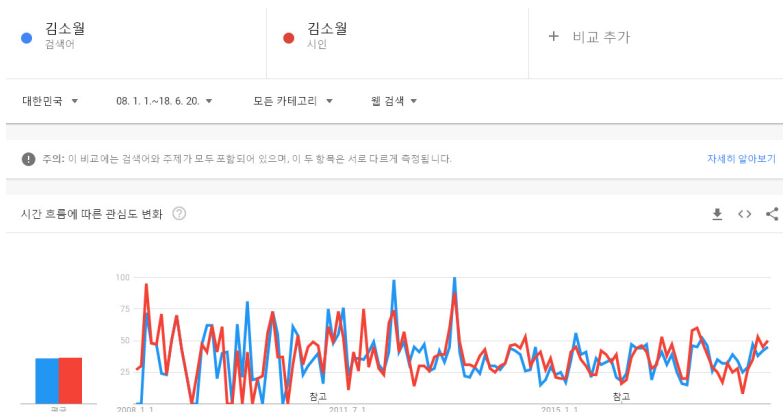


그림 2 김소월 탐색 조건

그림 3은 김수영의 관련 검색어이다. 파란색은 ‘김수영 검색어’로 탐색한 것이며, 붉은색은 ‘김수영 시인’으로 검색한 관련 검색어이다. 파란색의

관련 검색어를 살펴보면 ‘개그맨 김수영’, ‘김수영 작가’, ‘김수영 다이어트’, ‘이수영’ 등 시인 김수영과 관련 없는 검색어가 대다수이며, 오직 ‘4 김수영 폭포’만이 김수영의 시 「폭포」와 관련하여 검색된 결과임을 알 수 있다. 반면에 붉은색 ‘김수영 시인’ 관련 검색어를 살펴보면 ‘김수영 작가’, ‘이수영’을 제외하면 ‘김수영 풀’, ‘봄밤 김수영’, 신동엽 시인의 ‘껍데기는 가라’가 관련 검색어에 자리한다. 이것은 파란색 탐색보다는 붉은색 탐색이 시인 김수영과 더 밀접하게 관련된 검색어이자 검색량임을 알게 한다.

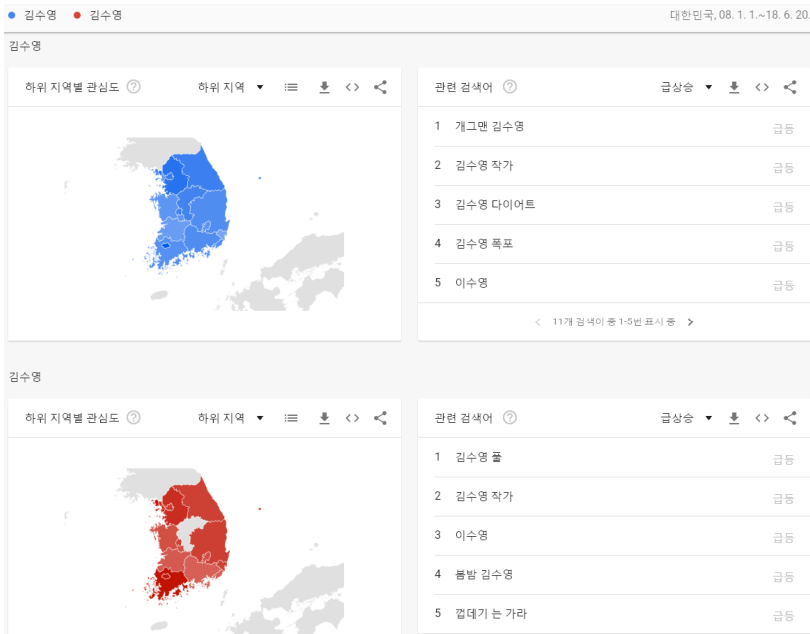


그림 3 김수영 탐색 조건 관련 검색어

파란색 ‘김수영 검색어’는 시인 김수영과 관련 없는 가변성의 요인이 다수 존재한다. 빅데이터 분석에서 가변성(Variability)이 증가하면 상대적으로 정확성(Veracity)은 감소하게 된다. 정확성은 데이터의 불확실성과 관련해 일정 유형의 데이터에 부여할 수 있는 신뢰 수준을 의미한다.(최도식,

2017, 269) 파란색 ‘김수영 검색어’로 탐색한 데이터를 활용할 경우 가변성의 증가로 시인 김수영에 대한 검색량의 신뢰 수준이 낮아지기 때문에 분고는 최소한도 내 가변성이 제거된 붉은색 ‘김수영 시인’에 대한 탐색을 데이터 분석 도구로 활용했다.

구글 트렌드는 검색어의 검색 횟수를 합산하여 최대 검색량을 절대값 100으로 하여 상대적 지표를 보여준다. 가장 높은 쿼리를 보이는 검색어를 100으로 설정하고 나머지 키워드의 쿼리 비중을 보여주는 것이다. 다시 말해, 해당 그래프, 관련 주제, 관련 검색어가 검색 쿼리수가 높고 낮은 검색 빈도가 가장 높은 검색어의 경우 100, 검색 빈도가 그 절반 정도인 검색어의 경우 50으로 나타난다. 그래서 총 100인의 대상 시인들에 대한 검색량 분석을 한 결과에서 꾸준한 검색 빈도가 없는 경우, 절대 검색 빈도 25가 두 번 이상 나타나지 않은 경우, 데이터 수집에 대한 신빙성이 부족한 2010년 이전 데이터에서 절대 검색 빈도가 반복된 경우 등에서 몇몇 데이터 신뢰에 대한 교란요인이 발생했다.(최도식, 268)

가령, 조지훈의 경우를 보면





그림 4 조지훈 검색량 및 관련 주제·검색어

조지훈에 대한 검색량은 2012년 1월 가장 많은 검색 빈도수를 보이고 있다. 그리고 그 외의 연도에서는 저조한 트랙량을 보인다. 그런데 조지훈의 관련 주제를 보면 조지훈 시인과 직접적으로 연관된 ‘시-주제’를 제외하면 개그맨 조지훈과 관련된 주제이다. 더욱이 검색 빈도수가 가장 많은 2012년 1월부터 4월까지의 뉴스 기사를 보면 개그맨 조지훈의 TV 광고 출연, 결혼, 쌍칼 통편집 등과 관련된 뉴스가 매일 수십 편이 올라와 있다. 그리고 그 이후의 연도에서는 시인 조지훈과 관련된 기사들이 간혹 올라와 있음을 확인할 수 있었다. 따라서 관련 주제는 개그맨 조지훈과 더 관련이 있으며, 관련 검색어의 경우 시인 조지훈과 관련되어 있음을 알 수 있다. 더욱이 조지훈의 대표시 「승무」, 「낙화」, 「고풍의상」 등이 관련 검색어에 순위를 올리고 있음을 확인할 수 있다.

조지훈에 대한 검색량 변화를 볼 때 2012년을 제외하면 25이하의 낮은 검색 빈도를 보이고 있다. 그리고 관련 주제 역시 시인 조지훈과 무관한 주제이다. 반면, 관련 검색어만이 시인 조지훈과 직접 관련된 검색어임을 알 수 있다. 다시 말해, 시간의 흐름에 따른 관심도 변화와 관련 주제는 데이터 분석의 교란요인이라 할 수 있다. 그러므로 검색량에서 꾸준한 검색 빈도가 없는 경우, 절대 검색 빈도 25가 두 번 이상 나타나지 않은 경우, 데이터 수집에 대한 신빙성이 부족한 2010년 이전 데이터에서 절대 검색 빈도가 반복된 경우 등의 시인은 비교 분석 대상에서 제외시켰다.

100인의 시인 중 다음의 충족조건을 갖춘 대상 시인은 많지 않았다. 충족조건을 갖춘 시인을 나열하면 다음과 같다.

구상, 기형도, 김소월, 김수영, 김영랑, 김용택, 김지하, 김춘수, 김현승, 나태주, 도종환, 류시화, 박목월, 박인환, 백석, 서정주, 안도현, 오장환, 유치환, 이상, 이성복, 정지용, 정호승, 한용운, 황동규(25인) + 고은, 윤동주 (교란요인을 감안한 추가 2인)

그런데 여기서 고은과 윤동주는 충족조건에서 제외되어야 하지만 ‘관련 주제와의 교란요인’, ‘절대 검색 빈도에 의한 교란요인’ 등 두 요인을 감안할 때 충족조건을 갖춘 것으로 파악하고 비교 분석 대상 시인에 포함시켰다(크리스천 리더, 2015, 128-151).

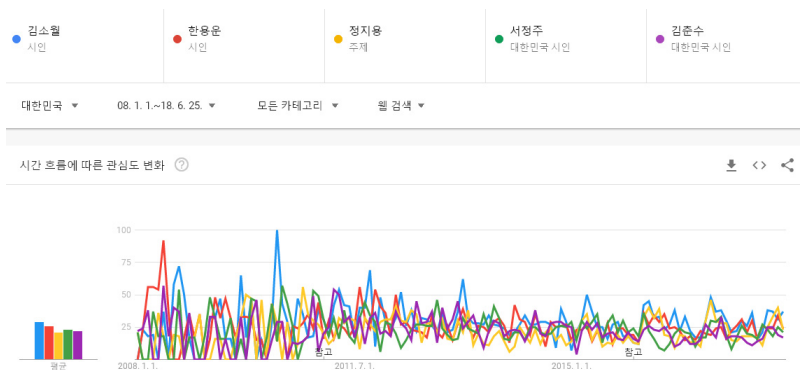


그림 5 작고시인 5인 검색량 비교 분석

본고는 비교 분석 대상의 기준으로 검색량 빈도가 꾸준하면서도 절대 검색 빈도가 25이상을 선회하는 김소월을 기준 시인으로 삼아 위의 조건에 충족된 시인들을 비교 분석하였다. 분석 결과를 보면 25인의 시인 중 김수영, 백석, 이상은 김소월보다 높은 검색량을 보였고 나머지 시인들은 김소월보다 낮은 검색량을 보였다.

검색량에서 김소월보다 낮은 검색량을 보이지만 거의 근사한 검색량을 보이는 시인은 한용운, 정지용, 서정주, 김춘수이다. 검색 빈도량을 보면 한용운→서정주→김춘수→정지용 순이다. 김소월이 가장 높은 검색 빈도량을 보이며, 관련 주제 및 검색어로 「진달래꽃」, 「초혼」, 「먼 후일」, 「산유화」, 「나의 집」 등의 시들이 관련 검색어로 검색된다. 한용운의 경우 「님의 침묵」, 「알 수 없어요」, 「복종」 등이 검색되며, 서정주의 경우 「국화 옆에서」, 「자화상」 등이 검색되었다. 김춘수의 경우 「꽃」이 관련 검색어로 검색된다. 그리고 정지용은 「향수」, 「호수」, 「춘설」 등이 관련 검색어로 검색된다. 이와 같이 작고 시인에 대한 분석 결과를 보면, 김소월, 한용운, 서정주, 김춘수, 정지용이 주요 검색 대상이 되고 있다.

한편, 김소월을 기준으로 기성 시인을 비교 분석해 보면 김소월에 비해 낮은 검색량을 보이며, 기성 시인 중 안도현, 도종환, 정호승이 거의 비슷한 검색량을 보인다. 이들 시인들은 작고 시인 중 검색량이 낮았던 정지용보다 낮은 검색량을 보이고 있다. 그래서 기성 시인만을 비교 분석했다.

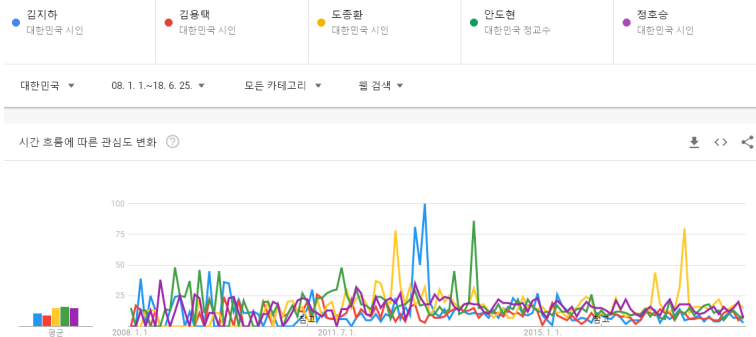


그림 6 기성시인 검색량 비교분석

기성 시인 중 가장 높은 검색량 빈도는 안도현이며, 도종환, 정호승이 그 뒤를 잇고 있으며, 김지하, 김용택 순으로 나타났다. 안도현의 관련 검색어는 「너에게 묻는다」, 「스며드는 것」, 「연탄 한 장」, 「간장 계장」, 「연탄재」 등이 검색되었으며, 도종환은 「접시꽃 당신」, 「담쟁이」 등의 시와 함께 「의원」, 「장관」 등이 검색되었다. 정호승은 「봄길」, 「수선화에게」, 「부치지 않은 편지」, 『내가 사랑하는 사람』, 『슬픔이 기쁨에게』 등의 작품과 시집이 검색되었다. 김지하는 「타는 목마름」, 「오적」 등의 시와 「변절」이 검색되었고, 김용택은 「섬진강」, 「참 좋은 당신」, 「봄날」, 「들국」 등의 시들이 검색되었다.

관련 주제 및 검색어를 볼 때 정호승의 작품이 다양하게 검색되고 있으며, 다량의 작품이 검색되고 있다. 그리고 안도현의 작품이 그 뒤를 잇고 있다. 반면에 도종환은 국회의원, 장관의 개인적 이력이 검색 빈도를 높이고 있으며, 김지하는 변절이라는 검색어에서 보이듯이 박근혜 전 대통령에 대한 지지로 한때 민중의 저항시인으로 손꼽혔던 시인에 대한 실망감이 반영된 결과로 분석된다.

교란요인으로 인해 분석 대상에서 배제한 윤동주와 고은에 대한 검색량을 추가적으로 살펴보고자 네이버의 검색어트렌드도 활용했다.

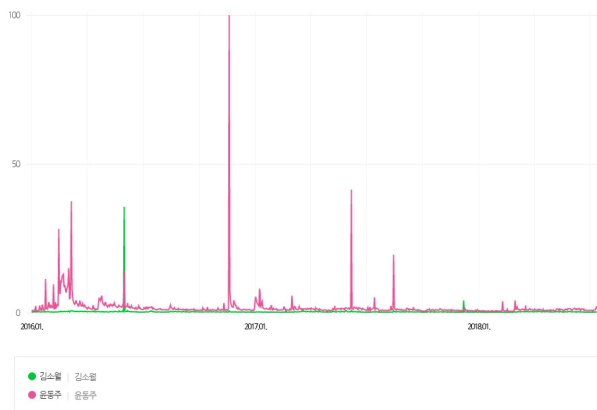


그림 7 김소월vs윤동주 검색 빈도 변화

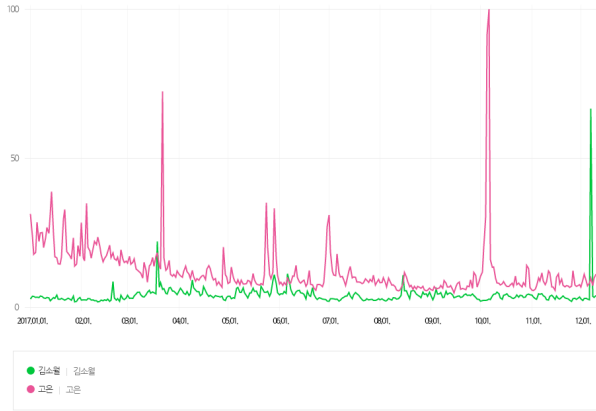


그림 8 김소월vs고은 연간 검색 빈도 변화

네이버 검색어트렌드에서도 구글 트렌드에서 보인 바와 유사한 결과도 출되었다. 김소월과 윤동주의 검색 빈도 변화에서 윤동주에 대한 검색 빈도가 높게 나타났으며, 김소월과 고은의 검색 빈도 변화에서도 고은의 검색량이 높게 나타났다. 그래서 다시 연간 검색 빈도 변화를 실행하였으며, 연간 검색 빈도에서도 동일한 결과가 나타났다. 그런데 고은의 경우 무작위로 1개월간 비교 검색했을 때 김소월이 더 높게 나타나고 있다. 더욱이 고은과 관련된 관련 주제 및 검색어를 보면, 걸그룹 멤버 레이샤의 본명 고은에 대한 “출사”, “레이샤 고은”이 관련 검색어 1·2위에 올라와 있고, 배우 김고은, 가수이자 배우인 별의 본명 김고은이 관련 검색어에 함께 검색되고 있다. 관련 검색어에 고은 시인의 시는 「눈길」, 「만인보」 뿐이다. 고은 시인의 실제 검색량은 뉴스 검색과 관련해 노벨상이 언급될 때 주로 언급되고 있으며, “최영미 시인”, “성폭력”이 관련 주제에 올라와 있음을 확인할 수 있다. 그리고 윤동주의 관련 주제 및 검색어는 영화 동주와 관련된 검색어가 다수 나타나고 있다는 점을 미루어 볼 때 윤동주 시인에 대한 검색만이 검색량에 포함된 것이 아니라 “송몽규”와 관련된 검색도 포함되어 있음을 알 수 있다. 그럼에도 불구하고 윤동주에 대한 관련 검색어는 그의 시 「서시」, 「별헤는 밤」, 「자화상」, 「십자가」, 「참회록」, 「고향」,

「쉽게 씌여진 시」, 「달을 쏘다」 등 대상 검색 시인 중 가장 많은 시가 관련 검색어에 올라와 있다. 이것은 윤희주에 대한 검색량이 단순히 영화 동주와 관련된 검색이 아니라 시인 윤희주에 대한 검색량으로 예측해 볼 수 있다(빅토르 마이어 원버거, 케네스 쿠키어, 2013).⁸

그리고 김소월보다 많은 검색량을 보인 김수영, 백석, 이상의 관련 주제 및 검색어를 살펴보면 김수영의 경우 개그맨 김수영과 관련된 검색이 다수 포함되어 있고, 백석의 경우 백석대학교와 일산의 백석동이 관련 검색어에 포함되어 있다. 이것은 순전히 시인 김수영과 백석에 대한 검색으로 볼 수 없다. 특히, 백석의 검색량은 과도하게 많은 검색 빈도를 보인다. 이러한 검색 빈도는 시인 백석에 대한 검색보다는 백석대학교와 백석동에 대한 검색이 주류 검색어로 예상된다. 실제 시인 백석에 대한 검색량은 작고 시인 5인 비교 분석에서 가장 낮은 검색량을 보인 정지용보다 더 낮은 검색량일 것으로 예측되는 바이다. 그리고 김수영의 경우도 「풀」, 「눈」, 「봄밤」 등의 시가 관련 검색어에서 검색되고 있다. 김수영의 경우 김소월과의 검색량 비교 분석을 했을 때 김소월보다는 낮은 검색량으로 한용운과 서정주 사이, 또는 서정주 정도의 검색량이 될 것으로 예측된다.

마지막으로 이상의 경우 작가 이상으로 김소월과의 비교 분석을 실행했다.

김소월에 비해 이상의 검색량이 높게 나타나고 있다. 그런데 관련 주제 및 검색어를 보면 김소월의 경우 그의 시가 대부분 관련 주제 및 검색어에 올라와있는 반면에 이상의 경우 「오감도」, 「이런 시」 외에도 소설 「날개」, 수필 「권태」가 관련 검색어에 올라와 있다. 이것은 이상을 시인으로만 검색한 결과라기보다는 소설가, 수필가로서의 검색량도 한몫하고 있음을 입증하는 것이다. 따라서 시인 이상의 검색량은 단순히 시인 이상으로만

8. 과거 스펠데이터를 통한 정확한 수치와 수량화가 중요하게 인식되었다면 지금은 정확한 수치보다는 들쭉날쭉messy한 데이터를 추출하고 처리하는 빅데이터가 중요하게 인식된다. 비록 두리몽실한 데이터, 들쭉날쭉한 데이터를 추출하고 처리해서 데이터를 변환하면 본래의 데이터가 다른 것으로 바뀌기 때문이다. 예컨대 영화 흥행 성적을 예측하기 위해 트위터 메시지에 담긴 정서를 분석하면 영화의 미래 예측 모델 및 감성분석, 평판분석이 이루어진다(빅토르 마이어 원버거-케네스 쿠키어, 2013, 65-69).

분석될 수 없는 한계가 있기에 소설가, 수필가로서 작가 이상으로 우리는 인식해야 할 것이다.

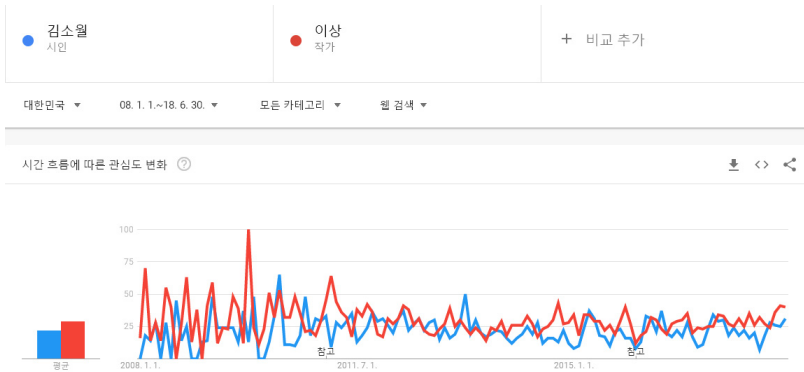
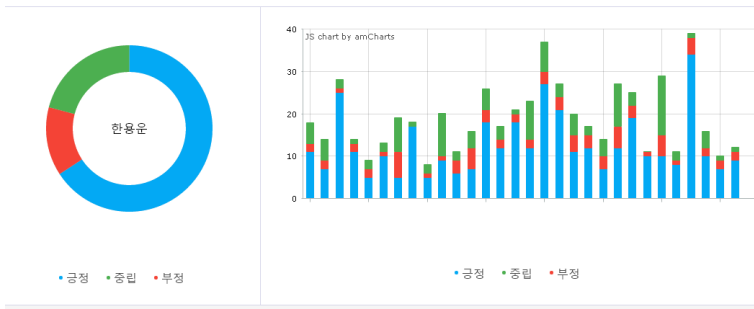


그림 9 김소월 vs 이상 검색량 비교

이상과 같이 구글 트렌드의 검색량 비교 분석을 통해 볼 때 작고 시인 중 대중이 가장 많이 검색하는 시인은 운동주로 예측되며, 김소월, 한용운 이상, 서정주·김수영, 정지용·김춘수로 예측된다. 그리고 기성 시인으로는 고은, 안도현, 도종환, 정호승, 김지하, 김용택으로 예측되지만 교란요인을 감안할 때 안도현, 정호승, 도종환·고은, 김지하·김용택으로 예측된다. 작품으로는 작고 시인의 경우 김춘수의 「꽃」, 운동주의 「서시」, 김소월의 「진달래꽃」, 정지용의 「향수」, 한용운의 「님의 침묵」, 김수영의 「풀」, 서정주의 「국화 옆에서」가 시인별 작품으로 검색되며, 기성 시인의 경우 안도현의 「너에게 묻는다」, 도종환의 「담쟁이」, 정호승의 「수선화에게」, 김지하의 「오적」, 안도현의 「스며드는 것」, 정호승의 「봄길」, 고은의 「눈길」, 김용택의 「섬진강」이 많이 검색되는 것으로 나온다.

이 결과를 보면 대중은 김소월에 대해 긍정적인 반응을 보인다. 이러한 대중의 반응은 그의 시에 대한 반응으로써 ‘사랑’, ‘좋아하다’, ‘사랑 받다’ 등과 같은 연애 감성을 불러오며, 그의 시와의 연관어에서 맺어지는 ‘그리움’, ‘아득함’ 등의 중립적인 반응이 추출되었다. 더욱이 인스타그램에는 ‘예전엔 밋치 몰랐서요’, 「금잔디」, 「못잊어」 등 구글 트렌드의 관련 검색어에서 볼 수 없었던 시들도 시집에서 캡처해 올라오고 있다. 그런데 김소월에 대한 부정 반응인 ‘눈물 흘리다’, ‘역겹다’ 등도 사실상 부정적인 반응으로 해석되기보다는 그의 시에 내재된 서러움의 정서와 「진달래꽃」의 ‘나보기가 역겨워’의 시어에서 추출된 긍정적인 시의 반응으로 보아야 한다.

한용운 금/부정 분석 상세



한용운 금/부정 추이

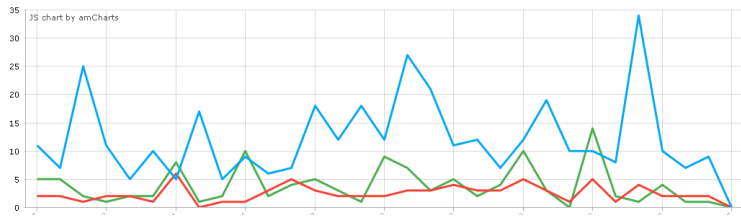


그림 11 한용운의 금/부정 분석 및 추이

그림 11의 분석 결과와 같이 한용운은 김소월보다 더 긍정 반응이 두드러진다. 연관 검색어의 감성 분석도 부정 반응보다 ‘사랑’, ‘행복’, ‘사랑하다’,

‘깨끗한’, ‘멋지다’, ‘믿다’ 등의 긍정 반응의 감성을 보인다. 중립 반응의 연관어 ‘말 없다’, ‘떠나다’, ‘고적하다’와 부정 반응의 연관어 ‘잊어버리다’는 김소월과 마찬가지로 그의 시 「님의 침묵」에 대한 반응으로 ‘침묵’, ‘이별’ 등과 연관해야 할 반응으로 분석된다. 이러한 분석을 볼 때 한용운은 김소월보다 대중에게 더 긍정적인 이미지로 다가와 있음을 알 수 있다.

위와 같은 데이터 수집 방법으로 대상 시인 13인에 대한 평판분석을 실행했다. 평판분석의 결과를 보면 연관어 언급량에서 긍정 반응이 가장 많은 시인은 도종환이었으며, 이어 김용택, 김춘수, 한용운, 정호승·안도현, 정지용, 김수영, 김소월, 고은, 서정주, 윤동주, 김지하로 나타났다. 그리고 긍정 반응 연관어로 ‘좋아하다’가 거의 대부분의 시인에게서 나타났으며, ‘사랑하다’·‘아름답다’가 그 뒤를 이었으며, ‘눈물흘리다’가 작고 시인에게서 많이 나타났다.

긍정반응 언급량이 가장 많은 도종환은 “내 문학과 내 인생의 희망은 좌절과 절망에서 시작했다. 여기서 인생을 끝낼 수 없다는 생각으로 일어섰고 결국 늦었지만, 인생의 꽃을 피웠다(맹석주, 2018).”, “국회의원을 하든, 장관을 하든 ‘시를 쓰는 사람’이라는 자기 정체성은 끝까지 지키려고 애를 씁니다. 시를 쓰는 사람이 세상을 바라보는 눈은 다름 아닌 ‘연민의 눈’입니다. 꽃 한 송이조차 연민의 눈으로 바라봤던 시인 시절처럼 문화·예술계의 다양한 관계자들을 만나고 다니는 지금도 그때의 그 마음을 잃지 않으려고 합니다(나윤석, 2018).” 등 트위터, 블로그, 뉴스 등에서 이품을 딛고 일어나는 희망과 사랑, 배려, 새로운 가치에 대한 긍정 감성 반응으로 읽혀진다. 그리고 김춘수의 경우 긍정 키워드로 “예쁜”에 대한 언급량이 가장 많이 나타나고 있다. 이것은 김춘수의 시 「꽃」에 대한 반응이며 “김춘수 시인의 ‘꽃’과 ‘국어사전에 숨은 예쁜 낱말’은 같다”는 트위터, “노작나무: ‘아침 이슬을 머금고 있어서 그런가 더 예쁜 거 같다.”는 블로그와 꽃 이미지미용 화장이미지 등의 인스타그램에 올라온 결과에 대한 반응이다. 다시 말해, 김춘수 시인에 대한 긍정 반응보다 그의 시 「꽃」에 대한 반응 내지는 식물 꽃에 대한 일반적인 반응과 미용에 대한 반응으로 읽혀진다. 이렇게

볼 때 김춘수의 긍정 반응 언급량이 높다고 하여 그 전체가 김춘수에 대한 긍정 반응이라고는 볼 수는 없을 것이다. 대개 긍정 반응에 대한 언급량은 기성 시인들에게서 더 많이 나타났다. 그런데 고은의 긍정 반응 언급량을 보면 이상 현상을 발견할 수 있다. 이상 현상은 평판분석의 오류, 곧 데이터 분석의 오류로 이어진다.

고은 시인의 긍정 반응 연관어를 보면 ‘고운’, ‘곱다’, ‘귀엽다’, ‘잘하다’, ‘좋다’, ‘후원받다’ 등이 있고 중립 반응 연관어에는 ‘새롭다’, ‘새로운’, ‘짧다’가 있으며, 부정 반응 연관어에는 ‘개소리’, ‘방해하다’가 있다. 긍정 반응 연관어와 중립 반응 연관어를 보면 고은 시인과 관련될 수 없는 키워드들이 있다. 이 연관어는 SNS 원문보기를 통해 걸그룹 레이샤 고은, 프로듀스48 나고은, 배우 고은아와 연관된 키워드임을 확인할 수 있었고, 20개의 언급량 중 시인 고은과 관련된 언급은 하나가 발견되었다. 이렇게 긍정 반응의 연관어 ‘고운’, ‘곱다’, ‘귀엽다’, ‘잘하다’와 중립 반응 연관어 ‘새롭다’, ‘새로운’, ‘짧다’를 제거하면 실제 남는 긍정 반응 연관어는 ‘좋다’, ‘후원받다’, ‘돈 벌다’가 있다. 그런데 ‘후원받다’의 경우 긍정반응에 대한 감성이라기보다는 부정반응에 대한 감성에 더 가깝다. 그리고 최근 연관어 언급량에서 상위에 랭크하고 있는 연관어가 ‘성폭력’, ‘폭력’, ‘성희롱’, ‘최영미’ 등이 있다. 이 연관어를 포함해 분석하면 고은 시인에 대한 대중적 감성은 긍정 반응이 우세하기보다 부정 반응이 더 우세한 것으로 예측된다.

한편, 작고 시인에게서는 긍정 반응과 더불어 부정 반응, 중립 반응이 혼재했다. 하지만 부정 반응과 중립 반응 비율이 높다고 하여 긍정 반응이 아닌 것은 아니다. 오피니언 마이닝을 활용한 평판분석은 여러 교란요인이 작동한다. 그래서 본 연구는 교란요인을 제거하고자 “SNS의 원문보기”를 통한 탐색어 여론에 대한 추가적 사실 확인을 적극 활용 했다. 그 결과 긍정 반응에 대한 신뢰도가 가장 높은 시인은 기성 시인인 정호승과 안도현으로 분석되었다.

정호승 감성 분석



정호승 긍/부정 분석 상세



그림 12 정호승 감성분석 및 긍/부정 비율

그림 12는 정호승의 감성분석이다. 긍정 반응에 대한 언급량 연관어로 ‘사랑’, ‘기쁨’, ‘아름다운’, ‘좋아하다’ 등이 있다. 이것은 당연히 정호승 시인의 시에 대한 반응이다. 반면에 부정 반응에 대한 언급량 연관어로 ‘분노’, ‘외로움’, ‘미안하다’, ‘고통’ 등이 있다. 그런데 “정호승 SNS 원문 보기”를 확인하면 정호승의 시 구절 “울지 마라. 분노 때문에 너 자신을 다치게 하지 마라. 네가 그를 사랑했다는 사실이 중요하다. 정호승, 「연인」, “미안하다 너를 사랑해서 미안하다… 정호승..!이런 시가 있었구나” 등에서처럼 분노, 외로움, 미안하다, 고통 등과 연관된 키워드를 확인할 수 있다. 이것은 정호승에 대한 부정 반응이라기보다는 그의 시에 대한 긍정 반응으로 분석되어야 한다. 그런데 이러한 반응은 안도현에게서도 공통적으로 나타났다.

안도현 감성 분석



안도현 긍/부정 분석 상세



그림 14 안도현 감성 분석 및 긍/부정 비율

안도현 SNS 원문 보기	
트위터 블로그 인스타그램 뉴스	
백인호	그대 이제 그만 마음아파해라 - 안도현: 별
글 읽어주는 츠나요시	살아가면서 가장 아름다운 일은 누군가의 배정이 되어주는 일이다. 별을 더욱 빛나게 해주는 까만하늘처럼, 꽃을 더욱 돋보이게 하는 무딘 땅처럼, 함께 하기에 더욱 아름다운 연어때처럼. 안도현, 연어
207*	오늘도 눈이 펑펑 쏟아진다. 흰 살 냄새가 난다. 그대 보고 싶은 내 마음 같다. 눈오는 날 / 안도현#자동
안도현 시 붓	길을 달려왔으나 정작 길을 데리고 오지는 못하였다는 자책이 물소리가 되어 발목을 묶는다 -안도현, '탁축도'
뿌앙3	연탄재 함부로 발로 차지 마라~너는 누구에게 한 번이라도 뜨거운 사람이었느냐너에게 묻는다-안도현]
hdnews9001	안도현의 「사랑」 http://blog.naver.com/hdnews9001

〈감성분석 결과에 따른 안도현 SNS 원문〉

안도현에 대한 부정 반응 언급량 연관어는 ‘아픔’, ‘슬픔’, ‘서글픈’ 등이 있고 중립 반응 언급량 연관어는 ‘스며들다’, ‘그립다’가 있다. 이 연관어를 볼 때 부정 반응 ‘아픔’은 안도현의 「별」에 대한 반응이며, ‘서글픈’은 ‘정작 길을 데리고 오지는 못하였다는 자책’(「탁축도」)에 대한 반응이며, 중립 반응 연관어 ‘스며들다’, ‘그립다’는 「너에게 묻는다」와 「눈오는 날」의 시 해석에 대한 반응으로 읽혀진다. 이렇게 볼 때 안도현에 대한 부정/중립 반응의 키워드는 긍정 반응으로 분석되어야 한다.

작고 시인 중에는 정지용과 윤동주의 부정/중립 반응에 주목할 필요가 있었다. 정지용의 부정 반응 연관어에는 ‘초라한’, ‘눈물 흘리다’ 등과 중립 반응 연관어에는 ‘따가운 햇살’이 있다. 이 부정/중립 반응도 정지용의 「향수」에서 추출된 반응임을 확인할 수 있다. 정지용 「향수」의 시어 중 ‘따가운 햇살’, ‘초라한’에 대한 부정 반응이며, 고향을 환기시키는 「향수」에 대한 ‘눈물’은 부정/중립 반응이 아니라 긍정 반응으로 분석되어야 한다. 따라서 정지용의 부정/중립 반응이 상당수 긍정 반응으로 분석되기에 정지용에 대한 긍정 반응은 상당히 높은 비율을 보일 것으로 예측된다.

윤동주는 중립 반응이 교란요인으로 작용한다. 중립 반응의 연관이 ‘충동 구매하다’는 윤동주 시인과 어떤 관계에 있는 반응일까?

윤동주 감성 분석

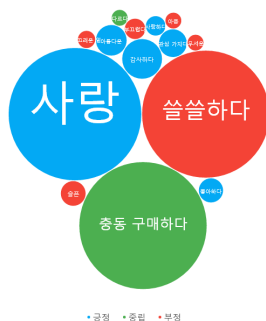


그림 15 윤동주 감성 분석 및 교란요인 에코 보틀

윤동주의 감성 분석에서 보이는 ‘충동 구매하다’는 최근 인터넷 쇼핑몰에서 판매되고 있는 윤동주의 「서시」가 새겨진 에코 보틀에 의한 연관어 언급량이라는 사실을 확인할 수 있었다. 그런데 에코 보틀 뿐만 아니라 윤동주 시가 새겨진 에코백, 머그컵 등과 그의 시집세트 등, 그의 시와 관련된 다양한 종류의 상품 판매에 의한 반응이었음을 확인할 수 있었다. 그런데 이 이상 현상은 이상에게서도 찾아진다. 이상의 「이런 시」가 대중에게 검색되는 이유도 이상의 「이런 시」를 에코 보틀로 제작 판매하고 있기 때문이다. 이러한 데이터의 결과를 추적해 볼 때 앞선 구글 트렌드 검색량에서 윤동주와 이상의 검색량이 많았던 이상 현상도 이들의 시가 에코 보틀, 에코백 등의 상품으로 제작 판매되고 있는 것과 무관하지 않을 것이다.

이상과 같이 평판분석에서 몇몇 교란요인을 감안할 때 긍정 반응이 높은 작고 시인은 김소월·한용운·정지용·윤동주이며, 기성 시인은 도종환·김용택·정호승·안도현으로 분석되었다. 그리고 김수영은 그의 아내와 관계된 사실(‘아내 때리다’)로 부정 반응이 있기는 하지만 긍정 반응이 다소 더 높은 것으로 예측된다. 서정주는 친일반민족행위자와 관련된 사실로

긍정/부정/중립 반응이 대등하게 나타나는 것으로 판단된다.⁹ 고은은 걸그룹 레이샤, 프로듀스 48 나고은과 관련된 긍정/중립 반응 연관어를 제거하면 부정 반응이 더 높은 것으로 예측된다. 그리고 김지하는 변절에 대한 이미지로 인해 중립/부정 반응이 높게 나타났다. 이러한 감성분석을 볼 때 기성 시인 중 고은과 김지하는 대중으로부터 점점 평판이 부정적 이미지로 나타나고 있음을 확인할 수 있다. 반면에 안도현, 정호승, 도종환, 김용택은 대중에게 긍정적 이미지가 점점 높아지는 것으로 예측된다. 따라서 제언하자면 한국 현대시의 세계화를 위한 번역 작업도 대중에게 긍정 평판이 높은 시인들의 시에 초점이 맞추어져야 할 것이다.

V. 맺음말

대중이 시인을 바라보는 평판은 그 시인이 작고 시인이든 기성 시인이든 언제든지 변할 수 있다. 시인의 개인사가 새로운 사실로 밝혀지거나 매체에 의한 이미지 전환이 이루어질 때 그들의 평판은 변한다. 그 평판의 한가운데에서 긍정적 이미지로 변하는 시인이 있는가하면 부정적 이미지로 변모하는 시인도 있다. 그렇기에 시인을 시로만 평가해야 한다는 주장에도 어느 정도 힘이 실린다. 하지만 대중으로부터 외면당한 시인은 그의 시도 외면당할 수밖에 없다.

그럼, 한국의 현대 시인들 중 대중에게 사랑받는 시인은 누가 있을까? 더욱이 한국 현대 시인들 중 누가 대중에게 자주 거명되며, 어떤 작품이 대중에게 사랑을 받고 있는 작품일까? 본 연구는 이 질문에 대한 답을 얻고자 빅데이터를 활용했다.

본 연구는 데이터 수집과 분석을 위해 빅데이터에 기반한 구글 트렌드와

9. 구글 트렌드에서 개별 시인 검색량을 살펴보면 친일반민족행위자로 서정주를 위시해 유치환, 모윤숙, 노천명, 김활란 등이 관련 검색어로 검색된다.

다음소프트의 썬트렌드, 소셜메트릭스를 분석 도구로 활용했다. 구글 트렌드를 활용해 시인별 검색량을 우선 분석한 후 김소월을 기준으로 각 시인을 비교 분석하였다. 그리고 다음소프트의 썬트렌드와 소셜메트릭스를 활용해 각 시인에 대한 감성분석, 긍정/부정/중립 반응 분석 및 추이, 연관어 등을 통해 평판분석을 시도했다.

분석 대상을 위한 대상 시인은 박재구, 강은교, 고정희, 고은, 구상, 기형도, 김경린, 김광규, 김광섭, 김규동, 김기린, 김기림, 김기택, 김남조, 김남주, 김동환, 김명인, 김소월, 김수영, 김승희, 김영랑, 김용택, 김종삼, 김지하, 김춘수, 김현승, 김혜순, 나태주, 나희덕, 노천명, 도종환, 류시화, 마종기, 모윤숙, 문덕수, 박남수, 박남준, 박노해, 박두진, 박목월, 박봉우, 박세영, 박세현, 박용래, 박인환, 박재삼, 백무산, 백석, 서정윤, 서정주, 송수권, 송옥, 신경림, 신동엽, 신석정, 신석초, 안도현, 오규원, 오세영, 오강환, 유안진, 유진오, 유치환, 윤동주, 이가림, 이근배, 이동주, 이상, 이상화, 이성교, 이성복, 이성부, 이수익, 이승훈, 이시영, 이용악, 이장희, 이해인, 이형기, 임화, 장정일, 전봉건, 정지용, 정진규, 정한모, 정현중, 정호승, 정희성, 조병화, 조정권, 조지훈, 천상병, 최두석, 최승자, 최승호, 하중오, 한용운, 한하운, 황동규, 황지우를 비롯해 총 100인을 선정했다.

구글 트렌드를 이용해 시인 100인의 검색량을 데이터로 수집하였다. 그 중 검색량에서 꾸준한 검색 빈도가 없는 경우, 절대 검색 빈도 25가 두 번 이상 반복되지 않는 경우, 데이터 수집에 대한 신빙성이 부족한 2010년 이전 데이터에서 절대 검색 빈도가 반복되는 경우의 시인을 비교 분석 대상에서 제외시켰다. 그리고 이를 충족시킨 25인의 시인과 교란 요인을 감안한 고은, 윤동주를 포함시켜서 27인을 비교 분석 대상으로 삼았고, 그 중 김소월을 기준 시인으로 설정해 비교 분석하였다.

분석 결과 작고 시인 검색 빈도량은 김소월, 한용운, 서정주, 김춘수, 정지용 순으로 나타났다. 가장 높은 검색 빈도량을 보인 김소월의 관련 주제 및 검색어로는 「진달래꽃」, 「초혼」, 「먼 후일」, 「산유화」, 「나의 집」 등의 시들이 검색되었다. 한용운의 경우 「님의 침묵」, 「알 수 없어요」,

「복종」 등이 검색되었으며, 서정주의 경우 「국화 옆에서」, 「자화상」 등이 검색되었다. 김춘수의 경우 「꽃」이 검색되었으며, 정지용은 「향수」, 「호수」, 「춘설」 등이 관련 검색어로 검색되었다. 그리고 작고 시인 중 백석은 여러 교란요인을 감안할 때 다섯 시인보다 낮은 검색량으로 예측되며, 이상은 관련 검색어 「날개」, 「권태」 등이 함께 검색되고 있는 점을 감안할 때 시인으로서의 검색량은 다섯 시인보다 적은 것으로 예측되었다. 윤동주는 「서시」, 「별헤는 밤」, 「자화상」, 「십자가」, 「참회록」, 「고향」, 「쉽게 씌여진 시」, 「달을 쏘다」 등 대상 검색 시인 중 가장 많은 시가 관련 검색어로 검색되고 있는 것으로 볼 때 김소월보다 높은 검색량 빈도를 보이는 것으로 예측했다. 마지막으로 김수영은 개그맨 김수영, 시나리오 작가 김수영과 관련된 검색이 다수 포함되어 있는 점을 미루어볼 때 김춘수, 정지용과 비슷한 검색량을 보이는 것으로 예측했다.

기성 시인 중 가장 높은 검색량 빈도를 보인 시인은 안도현이었으며, 도종환, 정호승이 그 뒤를 이었다. 그리고 김지하, 김용택 순으로 나타났으며, 『한국시학』, 『한국문학시대』가 노벨문학상 후보로 거명한 문덕수의 경우 검색량 빈도는 저조하였다. 한편, 검색량 빈도가 가장 높은 안도현의 관련 검색어는 「너에게 묻는다」, 「스며드는 것」, 「연탄 한 장」, 「간장 계장」, 「연탄재」 등이 검색되었으며, 도종환은 「접시꽃 당신」, 「담쟁이」 등의 시와 함께 「의원」, 「장관」 등이 검색되었다. 정호승은 「봄길」, 「수선화에게」, 「부치지 않은 편지」, 『내가 사랑하는 사람』, 『슬픔이 기쁨에게』 등의 작품과 시집이 검색되었다. 김지하는 「타는 목마름」, 「오적」 등의 시와 「변질」이 검색되었고, 김용택은 「섬진강」, 「참 좋은 당신」, 「봄날」, 「들국」 등 그의 다양한 시들이 검색되었지만 의외로 다섯 시인 중 가장 낮은 검색량 빈도를 보였다. 그리고 고은의 검색량은 김소월보다 상당히 높게 나타났지만 걸그룹 레이샤의 리더 고은, 배우 김고은, 프로듀스48의 나고은 등에 대한 검색량이 포함된 것으로 볼 수 있다. 이런 교란요인을 감안할 때 고은의 검색량은 김지하보다는 높은 도종환, 정호승 정도의 검색량으로 예측했다.

평판분석에서 연관어 언급량의 긍정 반응은 도종환, 김용택, 김춘수, 한용운, 정호승·안도현, 정지용, 김수영, 김소월, 고은, 서정주, 윤동주, 김지하 순으로 나타났다. 그런데 부정 반응이 시인의 시 이미지에 대한 반응으로 판단되는 김소월, 한용운, 정호승, 안도현, 정지용의 경우 부정 반응을 긍정 반응으로 해석해야 하며, 김춘수의 긍정 반응의 일부는 식물, 미용과 관련된 이미지로 해석되며, 고은의 긍정 반응은 구글 검색량 관련 검색어 레이샤, 김고은, 나고은 등과 연관시켰을 때 긍정 반응보다 부정 반응이 더 우세한 것으로 예측되었다. 그리고 윤동주의 중립 반응은 인터넷 쇼핑몰과 관련된 교란요인으로 해석되었고 구글 트렌트에서 윤동주의 검색량이 과도하게 많은 이유가 인터넷 쇼핑몰의 상품 판매와 무관하지 않음을 확인할 수 있었다.

평판분석에서 교란요인을 감안할 때 긍정 반응이 높은 작고 시인은 김소월·한용운·정지용·윤동주이며, 기성 시인은 도종환·김용택·정호승·안도현으로 분석되었다. 그리고 김수영은 부정 반응이 있기는 하지만 긍정 반응이 다소 더 높은 것으로 예측되며, 서정주는 친일반민족행위자와 관련된 사실로 긍정/부정/중립 반응이 대등하게 나타나는 것으로 판단된다. 고은은 교란요인의 긍정/중립 반응 연관어를 제거하면 부정 반응이 더 높은 것으로 예측되며, 김지하는 변절에 대한 이미지로 인해 중립/부정 반응이 높게 나타났다. 이러한 감성분석을 볼 때 기성 시인 중 고은과 김지하는 대중으로부터 점점 평판이 부정적 이미지로 나타나고 있음을 확인할 수 있다. 반면에 안도현, 정호승, 도종환, 김용택은 대중적 인지도가 점점 높아질 뿐 아니라 긍정적 이미지도 점점 높아지는 것으로 예측된다.

참고문헌

- 강만모·김상락·박상무(2012), 「빅 데이터의 분석과 활용」, 『정보과학회지』 30(6), 26쪽.
- 권혁웅·전윤호(2002), 「시인·평론가 100인이 뽑은 ‘10명의 시인’ 선정 경위」, 『시인세계』 (가을호), 27-33쪽.
- 김기봉(2016), 「빅데이터의 도전과 인문학의 응전」, 『시민인문학』 30, 9-37쪽.
- 김병선(2013), 「현대시인의 문체적 지문을 찾아서」, 『국어국문학』 143, 153-188쪽.
- _____(2000), 「현대시의 계량적 문체 연구 시론」, 『국어국문학회』 전국학술대회 발표논문집, 10-27쪽.
- 김연권(2016), 「빅데이터 시대에 인문학의 역할과 과제」, 『시민인문학』 30, 38-69쪽.
- 김용재(2017), 「문덕수 시인 『우체부』의 국제적 위상」, 『한국문학시대』 (여름호), 26쪽.
- 김학동 외(2005-2007), 『한국 전후 문제시인 연구 1-7』, 예림기획.
- 김호연(2016), 「디지털 기술과 인문학의 융합」, 『시민인문학』 30, 70-95쪽.
- 김희전·조현양(2010), 「저자동시인용분석과 저자서지결합분석에 의한 지적 구조 분석」, 『정보관리학회』 27(3), 283-306쪽.
- 나윤석, 『매일경제』, 2018.7.15일자, “도중환 “연민의 눈으로 세상보는 시인의 정체성 잃지 않을 것””.
- 류찬열(2008), 「1970년대 한국 현대시의 계보와 유형 연구를 위한 시론」, 『우리문학연구』 24, 203-225쪽.
- 맹석주(2018), 『노컷뉴스』, 2018.6.4일자, “도중환 장관 일신여고에서 특강”.
- 박자현·송민(2013), 「토픽모델링을 활용한 국내 문헌정보학 연구동향 분석」, 『정보관리학회지』 30(1), 7-32쪽.
- 박진균·김택윤·송민(2017), 「텍스트마이닝을 이용한 윤동주 연구의 개체계량학적 분석」, 『한국비블리아학회지』 28(1), 191-207쪽.
- 유강하(2016), 「빅데이터와 사물인터넷 시대의 비판적 해석과 인문학적 상상력」, 『시민인문학』 30, 96-116쪽.
- 임현순(2008), 「독자들의 한국현대시 수용과 인식의 특성」, 『한국시학연구』 22, 320-321쪽.
- 정과리(2016), 「빅데이터의 문학적 활용과 빅데이터에 대한 문학적 성찰」, 『시민인문학』 30, 117-136쪽.
- 정유진·손민지, 『중앙일보』, 2016.10.31일자 “한국이 노벨문학상 못 받는

진 번역 탓?”.

조용호, 『세계일보』, 2005.10.13일자, “한국, 노벨문학상에 “끝없는 짝사랑””.
최도식(2017), 「빅데이터 분석 교육의 문제점과 개선 방안」, 『한국융합학회논문지』 8(12), 269쪽.

조선일보편집부, 『조선일보』, 2014.7.1일자, “문덕수 前 문예진흥원장 시집 출간”.

조선일보편집부, 『조선일보』, 1998.7.31일자, “문학 50년, 건국 이후 뛰어난 시인 49인”.

빅토르 마이어 쉰버거·케네스 쿠키어(2013), 『빅데이터가 만드는 세상』, 이지연 역, 21세기북스, 65-69쪽.

임마누엘 페스트라이쉬(2013), 『한국인만 모르는 다른 대한민국』, 21세기북스.
크리스천 러더(2015), 이가영 역, 『빅데이터 인간을 해석하다』, 도서출판 다룬, 128-151쪽.

McKinsey (2011.6), *McKinsey Global Institute-Big data : The next frontier for innovation-competition and productivity*, McKinsey & Company, p.1.

『구글 트렌드』, 구글, <https://trends.google.com>.

『썸트렌드』, 다음소프트, <http://www.some.co.kr>.

『소셜 메트릭스』, 다음소프트, <http://www.socialmetrics.co.kr/socialSearch.html>.

국문초록

한국 현대 시인들 중 누가 대중에게 자주 검색될까? 어떤 시가 대중에게 많이 검색될까? 이 질문에 대한 답을 얻고자 최근 각광 받고 있는 빅데이터를 활용했다.

본 연구의 데이터 수집과 분석은 구글 트렌드와 썸트렌드, 소셜메트릭스를 이용했다. 분석 절차는 100인의 시인에 대해 구글 트렌드로 검색량을 탐색했으며, 검색 빈도가 높은 27인의 시인을 김소월과 비교 분석했다. 그리고 27인의 시인 중 13인의 시인을 대상으로 썸트렌드, 소셜 메트릭스를 이용해 감성분석, 긍/부정 반응 등의 평판을 분석했다.

검색량 분석 결과 작고 시인 검색량은 김소월, 한용운, 서정주, 김춘수, 정지용이 높았다. 김소월의 관련 검색어로는 「진달래꽃」·「초혼」·「먼 후일」·「산유화」·「나의 집」, 한용운은 「님의 침묵」·「알 수 없어요」·「복종」, 서정주는 「국화 앞에서」·「자화상」, 김춘수는 「꽃」, 정지용은 「향수」·「호수」·「춘설」이 검색되었다. 그런데 김소월보다 높은 검색량에는 윤동주, 김수영이 탐색되었다. 윤동주의 관련 검색어는 「서시」·「별헤는 밤」·「자화상」·「십자가」·「참회록」·「고향」·「쉽게 씌여진 시」·「달을 쏘다」가 검색되었고 김수영은 개그맨 김수영, 시나리오 작가 김수영과 관련된 검색이 다수 포함된 점을 감안할 때 김춘수, 정지용과 비슷한 검색량으로 예측했다.

기성 시인 검색량은 안도현, 도종환, 정호승, 김지하, 김용택 순이었다. 안도현은 「너에게 묻는다」·「스며드는 것」·「연탄 한 장」과 “간장 계장”·“연탄재”, 도종환은 「접시꽃 당신」·「담쟁이」와 “의원”·“장관”, 정호승은 「봄길」·「수선화에게」·「부치지 않은 편지」·『내가 사랑하는 사람』·『슬픔이 기쁨에게』, 김지하는 「타는 목마름」·「오적」과 “변절”, 김용택은 「섬진강」, 「참 좋은 당신」, 「봄날」, 「들국」이 검색되었다. 그런데 고은의 검색량은 김소월보다 높게 나타났지만 걸그룹 레이샤의 리더 고은, 배우 김고은 등 교란요인이 있기에 이를 감안할 때 김지하보다는 높은 도종환, 정호승

정도의 검색량으로 예측되었다.

평판분석에서 연관어 언급량의 긍정 반응은 도종환, 김용택, 김춘수, 한용운, 정호승·안도현, 정지용, 김수영, 김소월, 고은, 서정주, 윤동주, 김지하 순으로 나타났다. 김소월, 한용운, 정호승, 안도현, 정지용의 부정 반응은 긍정 반응으로 해석되며, 김춘수의 긍정 반응은 순전한 긍정 반응이 아니었으며, 고은의 긍정 반응은 레이샤, 김고은 등과 연관되므로 부정 반응이 더 우세한 것으로 예측되었다. 그리고 윤동주의 중립 반응은 인터넷 쇼핑몰과 관련된 교란요인으로 해석되었고 구글 트렌드에서 윤동주의 검색량이 과도하게 많은 이유가 인터넷 쇼핑몰의 상품 판매와 무관하지 않음을 확인할 수 있었다.

교란요인을 감안한 긍정 평판 우세는 김소월·한용운·윤동주·정지용과 도종환·김용택·정호승·안도현으로 분석되었다. 고은은 교란요인의 긍정/중립 반응 연관어를 감안하면 부정 평판이 더 높은 것으로 예측되며, 김지하는 변절에 대한 이미지로 인해 중립/부정 반응이 높게 나타났다. 기성 시인 중 고은과 김지하는 대중으로부터 점점 부정적 평판을 받고 있으며, 안도현, 정호승, 도종환, 김용택은 점점 긍정적 평판이 높아지는 것으로 예측된다. 이에 제언하자면 한국 현대시의 세계화를 위한 번역 작업도 안도현, 정호승, 도종환, 김용택 등의 시인에게 초점이 맞추어졌으면 하는 바람이다.

키워드: 현대 시인, 빅데이터, 구글 트렌드, 검색량, 평판분석, 김소월

Abstract

Analysis of Public Awareness and Opinion Mining of Modern Poet through Big Data

Do-sik Choi(Kangwon National Univ.)

Which of the modern poets of Korea will be frequently found in the public? Which poems will be found in the public? To answer this question, they used Big Data, which has recently been in the spotlight. The study's data collection and analysis used Google Trends, thumb trends, and social metrics. The analysis procedure searched for 100 poets with Google Trends. 27 people with high frequency of search were compared. The study had a reputation analysis of 13 of the 27 poets. The first analysis of the study showed that Kim So-wol, Han Yong-woon, Seo Jung-joo, Kim Chun-soo and Jung Ji-yong had a high number of searches. The relevant search result for the study was Kim So-wol's "Azaleas", "Invocation of the spirits of the dead", "Some distant day", "A mountain flower", "My house" and Han Yong-woon's "Your silence", "I don't know", "Obedience", Seo Jeong-joo's "By Chrysanthemum", "Self-portrait", Kim Chun-Soo is "Flower", Jung Ji-young is "Fragrance", "Lake", "Springtime". Yoon Dong-joo is "Prologue poem", "Starry Night", "Self-portrait", "Cross", "Confession", "Hometown", "Written easily", "Shoot the Moon". The Poet Kim Soo-young had many searches for the comedian Kim Soo-young.

The living poets are the order of An Do-hyeun, Do Jong-hwan, Jung Ho-seung, Kim Ji-ha, and Kim Yong-taek. The search word of An Do-hyeon discovered as "Ask you", "Permeating there", "A piece of briquette" and 'a used briquet', 'briquette' of poetic diction. Do Jong-hwan discovered as "You, the Hollyhock", "An ivy" and 'senator', 'minister', poet Jung Ho-seung discovered as "Spring road", "To Daffodil", "Unbearable letter". And Kim Ji-ha searched out "Burning thirst", Kim Yong-taek searched out "Seomjin River", "You are so good", "Spring

day" and "A wild chrysanthemum". However, the poet Ko eun was more searched by Ko eun, the leader of the girl group.

In the analysis of Opinion Mining, affirmative responses of related words are in order of Do Jong-hwan, Kim Yong-taek, Kim Chun-Soo, Han Yong-woon, Jung Ho-seung, An Do-hyeon, Jung Ji-young, Kim So-wol, Seo Jeong-joo, Yoon Dong-joo, Kim Ji-ha.

Yoon Dong-joo's neutral reaction was interpreted as disturbance factor related to internet shopping mall. Yoon Dong-joo's search volume in Google Trent was excessive due to sales of Internet shopping malls. In the analysis of Opinion Mining, the highest order of positive response was analyzed by Kim So-wol, Han Yong-woon, Yoon Dong-joo, Jung Ji-young, Do Jong-hwan, Kim Yong-taek, Jung Ho-seung, An Do-hyeon. Ko eun poet and Kim Ji-ha poet have received increasingly negative reputation from the public. An Do-hyeon, Jung Ho-seung, Do Jong-hwan and Kim Yong-taek are expected to increase their positive reputation gradually. Therefore, the translation work for the globalization of modern Korean poetry should be focused on the poems of An Do-hyeon, Jung Ho-seung, Do Jong-hwan and Kim Yong-taek.

Keywords: Modern Poet, Big-Data, Google Trent, Keyword Search Volume, Opinion Mining, Kim So-wol

Received :	12 July	2018
Reviewed :	1 August	2018
Accepted :	10 August	2018