

텍스트마이닝 기반의 효율적인 장소 브랜드 이미지 강도 측정 방법*

최석재

경희대학교 후마니타스 빅데이터 연구센터
(sjchoi@khu.ac.kr)

전종식

경희대학교 후마니타스 빅데이터 연구센터
(jseon@khu.ac.kr)

비스웬스 수브르더

경희대학교 일반대학원 경영학과
(subhongiku.09@khu.ac.kr)

권오병

경희대학교 경영학과
(obkwon@khu.ac.kr)

.....

장소 브랜드는 특정 장소에 대한 의미 부여를 통해 장소성의 정체성 및 공동가치를 생성하며 가치 창출을 하는데 중요한 활동이며, 장소 브랜드에 대한 이미지 파악을 통해 이루어진다. 이에 마케팅, 건축학, 도시건설학 등 여러 분야에서는 인상적인 장소 브랜드의 이미지를 구축하기 위하여 많은 노력을 기울이고 있다. 하지만 설문조사를 포함한 대면조사 방법은 대부분 주관적인 작업이며 측정에 많은 인력 또는 고도의 전문 인력이 소요되어 고비용을 발생시키므로 보다 객관적이면서도 비용효과적인 브랜드 이미지 조사 방법이 필요하다. 이에 본 논문은 텍스트마이닝을 통하여 장소 브랜드의 이미지 강도를 객관적이고 저비용으로 얻는 방법을 찾는 것을 목적으로 한다. 제안하는 방법은 장소 브랜드 이미지를 구성하고 있는 요인과 그 키워드들을 관련 웹문서에서 추출하며, 추출된 정보를 통해 특정 장소의 브랜드 이미지 강도를 측정하는 방법이다. 성능은 안홀트 방법에서 평가에 사용하는 전세계 50개 도시 이미지 인덱스 순위와의 일치도로 검증하였다. 성능 비교를 위해 임의로 순위를 매기는 방법, 안홀트의 설문방식대로 일반인이 평가하는 방법, 본 논문의 방법을 사용하되 안홀트의 방법으로 학습한 것으로 유의한 것으로 추정되는 평가 항목만을 반영하는 방법과 비교하였다. 그 결과 제안된 방법론은 정확성, 비용효율성, 적시성, 확장성, 그리고 신뢰성 측면에서 우수함을 보일 수 있었다. 따라서 본 연구에서 제안한 방법론은 안홀트 방식에 상호 보완적으로 사용될 수 있을 것이다. 향후에는 장소 브랜드 이미지를 형성하는 속성 별로 등장횟수를 계산 한 후에 장소 브랜드에 대한 태도, 연상, 그리고 브랜드 자산과의 인과관계를 자동으로 파악할 수 있는 부분까지 구현하고 실증적 실험을 할 예정이다.

주제어 : 텍스트마이닝, 브랜드 이미지, 장소 브랜드, 주제어 분석, 안홀트 인덱스

.....

논문접수일 : 2015년 6월 1일 논문수정일 : 2015년 6월 15일 게재확정일 : 2015년 6월 16일
투고유형 : 국문급행 교신저자 : 권오병

1. 서론

브랜드 이미지란 브랜드와 관련하여 사람들이 연상하는 모든 것의 복합적인 이미지이며 (Newman, 1957), 의미를 지니고 조직화된 연상

들의 집합으로 (Aaker, 1990), 또는 브랜드 속성, 편익, 태도 등의 다양한 브랜드 연상을 통해 반영된 브랜드에 대한 인식으로 정의된다(Keller, 1993). 따라서 브랜드 이미지는 기능적, 경제적, 사회적 등의 여러 차원을 가지고 있다. 즉, 장소

* 이 논문은 2014년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-20141823). 또한 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [R0126-15-1007, Curation commerce based global open market system development for personal happiness enhancement]

브랜드는 특정 장소(예: 도시, 관광지 등)의 자연 환경이나 역사적인 특성, 문화적인 매력, 서비스, 고유한 정체성(identity) 등을 인식하게 하면서 다른 장소와 차별화된 본질적 개념(concept)을 통해 주민, 기업인, 방문객 등 고객 및 잠재고객들이 부가적으로 가치를 느끼게 하는 장소의 이름, 장소의 심볼, 캐릭터, 슬로건, 명소의 이름, 각종 하위 브랜드(예: 축제 브랜드, 공동브랜드, 특산물브랜드 등)들이 다 포함된 총체적인 상징체계이다 (Gu and Lee, 2008).

브랜드 이미지는 도시 등 특정 장소에 대한 평가나 공간 마케팅을 수립하는 데 중요하고 유용한 개념이다. 장소 자산의 의미 부여를 통해 장소 브랜드를 포지셔닝하고, 이를 통해 장소 자산 활동으로서의 상징적 사건 교류와 환경으로서의 물리적 공간 체험과 같은 장소 자산 지향 활동을 통해 의미, 활동, 물리적 환경이라는 세 가지 차원으로 설명되는 장소 브랜드 이미지를 조성한다(Lee, 2014). 이렇게 형성된 이미지를 통해 유형적 무형적 가치 창출을 하게 되는 것이다.

그런데 장소 브랜딩을 위한 이미지 분석은 일반적으로 문헌연구 또는 면담을 통한 전문가의 조언이나 일반인을 대상으로 하는 설문 조사의 방법으로 이루어져 왔다. 장소 이미지 획득을 위해 사용되는 대표적인 설문 방식으로는 자유연상법 등이 있다(Lee and Lim, 2005). 이는 일반인들로 하여금 특정 장소의 명칭을 들을 때 연상되는 것을 기록하게 하는 것이다. 그러나 이 방법은 정적(static)이어서 적시적이고 기민한 분석이 어려우며, 비록 가능할지라도 적지 않은 비용이 발생하게 된다. 또한 상품화되는 특정 장소(도시, 관광지 등)와 관련되는 것은 소비자가 직접 체험 후 얻어지는 경험재(experience goods)의 성격을 가지는데 경험을 하지 않은 일반인의 진술은 신

뢰하기 어려운 것이 된다. 따라서 이러한 문제를 해결하기 위해서는 우선 진술자들이 장소에 대한 경험자여야 하고, 그러한 진술자들의 진술과 평가를 대량으로 수집하고 분석 방법을 체계화하는 방법이 필요하다. 대신 경험을 한 진술자들에 의하여 기록된 특정 장소에 대한 내용을 널리 확보하여 이미지 분석에 활용할 수 있다면 장소 마케팅을 수행하는 전문가들에게 장소와 관련된 경험 정보는 물론, 아이디어 창출을 위한 직관을 제공할 수 있을 것이다. 특히 최근 다양한 분야에서 텍스트 마이닝 기법을 활용하여 정량적 분석을 하는 연구(Kim et al., 2015; Yi et al., 2015 등)들이 활발하고 성공적으로 이루어지고 있는 점을 고려하면 기존의 방대한 시간과 자원을 들여 사람들의 의견을 수작업으로(manually) 획득 하였던 기존의 작업들은 자동화 방안 가능성을 전면적으로 검토할 필요가 있다. 그러나 아직까지 장소 브랜드 이미지를 자동으로 추출하는 정교화된 텍스트마이닝 방법은 소개되지 않았다.

따라서 본 논문의 목적은 텍스트마이닝 기법을 기반으로 동적(dynamic)이고 적시적이며, 비용효율적인 장소브랜드 이미지를 추출하는 방법론을 제시하는 것이다. 제안한 방법론의 성능을 검증하기 위해서 설문조사법에 의한 도시 브랜드 이미지 평가로 가장 널리 알려진 안홀트(Anholt)의 방법에 의하여 도출된 결과와 비교한다. 지역 브랜드 이미지 평가에 관련하여 보다 체계적이고 심도 있는 조사된 것을 채택하고자 하였으나 이에 대한 기존의 평가가 주관적이고 소규모로 이루어진 것이 많고 공신력 측면에 있어서 부족한 것이 많아 본 연구 논문에서는 도시 브랜드 이미지에 초점을 두어 수행하였다.

본 논문의 구성은 다음과 같다. 먼저 제2장에서는 본 논문의 주제인 장소브랜드 이미지를 평

가하는 기존의 방법들에 대한 문헌을 정리하여 기술하였다. 제3장과 제4장은 각각 본 논문에서 제안하는 도시 브랜드 이미지 평가 방법론을 안홀트 방법론을 기반으로 하여 소개하고, 평가하기 위한 방법론과 평가 결과를 보였으며, 마지막으로 제5장에서 제안 방법론의 시사점과 추후 연구방향에 대해 논의하였다.

2. 장소브랜드 이미지 평가

브랜드 이미지는 브랜드 자산(brand equity)에 영향을 미치고 이는 곧 성공적인 경영에 관련되므로 브랜드 이미지 및 브랜드 자산은 학계뿐 아니라 경영 실무자(practitioners)들에게도 높은 관심의 대상이 되어 왔다 (Cuneo, 1999; Faircloth et al., 2001). 그 동안 장소 브랜드의 이미지에 영향을 미치는 전략적 요소들에 대한 연구가 진행되어 왔다. 장소 브랜딩 전략 요소란 장소 브랜딩을 실시할 경우 반드시 고려해야 할 요소를 말한다. 선행 연구에 나타난 장소 브랜딩의 전략 요소를 키워드로 추출한 것은 <Table 1>과 같다.

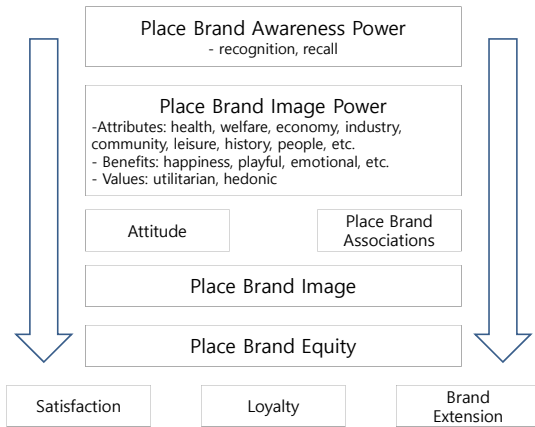
이러한 장소 브랜드에 대한 연구는 장소 마케팅

의 이유에 따라 더 세분화될 수 있다. 예를 들어 장소로서의 관광지에 대한 장소 브랜드는 아래와 같이 관광지로서의 경쟁력을 평가하기 위해서 관광 경험의 질(Quality of Experience), 관광지에서 만나게 되는 브랜드 접촉점으로서의 현지인, 만족할 만한 경험을 보장하기 위한 전문인력, 장소 브랜드로서의 관광지가 갖는 환경적, 물리적 요소, 관광 활동으로 관광객이 방문지에서 관광객이 경험하는 행동적 요소 그리고 관광산업의 관련 문화 등 크게 다섯 가지 측면에서 살펴 볼 수 있다.

한편 브랜드의 이미지의 강도를 측정하기 위하여 고안된 것이 브랜드 파워 모형(Brand power model)이다(Na et al. 1999). <Figure 1>과 같이 장소라는 브랜드에 대한 브랜드 파워 모형은 브랜드 인식(brand awareness)의 파워가 브랜드 이미지 파워에 영향을 주고, 이는 브랜드에 대한 태도 및 연상의 형태로 브랜드 이미지에 영향을 주게 되며, 결국 브랜드 자산(brand equity)에까지도 영향을 준다고 본다. 여기서 브랜드 이미지 파워는 브랜드의 속성과 그 속성이 제공할 것으로 보이는 이익(benefit)을 거쳐 브랜드 가치(value)가 형성되어 이루어진다고 보았다. 이를 위해 브랜드

<Table 1> Previous researches on place branding

Researcher	Branding Element
Anholt (2006)	presence, place, prerequisite, people, pulse, potential
Aaker (1995)	brand awareness, perceived quality, brand loyalty, brand associations
Keller (1993)	brand awareness, brand image
Lee (2005)	Awareness, preference, value for visit, uniqueness, popularity, price premium
Park and Chang (2009)	symbolic identity, associations, story, landmark, creativity, substantial values, vision, evolution, global competitiveness, overall view
Chang et al. (2010)	identity, quality, potential, uniqueness, value for visit
Lee and Lim (2011)	symbolic image, value creating, storytelling, sustainability, competitiveness
Lee (2014)	economical value, relational value, creating value, recycling value



〈Figure 1〉 Model of place brand Image power

드 이미지 브랜드 파워 모형은 브랜드 자산 측정에 있어 복수 개의 속성을 고려할 수 있도록 한다. 그리고 브랜드 파워 모형에 의하면 브랜드 이미지에 대한 파워(Brand Image Power, BIP)는 다음과 같다.

$$BIP_{ij} = \sum_{\forall k} W_{ik} X_{jk}$$

단 여기서,

BIP_{ij} = 의사결정자 i에게 있어서 대안 j의 전체적인 브랜드 파워

k = 속성, 이익, 가치 함수 ($k = 1,2,3,\dots,K$),
 X_{jk} = k번째 속성, 이익, 가치 함수에 의하여 계산된 대안 j의 점수
 W_{ik} = 의사결정자 i에게 있어서 k번째 속성, 이익, 가치 함수가 가지는 상대적 비중

브랜드 파워 모형은 물리적 장소뿐 아니라 웹사이트와 같은 가상 공간 마케팅(Na, 2005)에도 적용된 바 있다. 하지만 일반적인 브랜드에 비하여 장소 브랜드가 가지는 특성을 잘 반영하지는 못했다. 또한 브랜드란 의사소통자(communicator)에 의해 개념화되는 것인 만큼(Hankinson, 2004), 의사소통자가 진술한 내용을 확보하여 분석하는 것이 브랜드 이미지 형성 연구에 중요함에도 불구하고 이러한 내용 분석 방법을 실시하는 연구는 그리 많지 않다. 또한 브랜드는 이름, 심볼, 디자인 결과물 및 이들의 합작품으로 이해되므로(American Marketing Association, 1960), 이미지 개념은 다양하고 비정형적인 내용으로 표현되지만 이러한 부분을 고려한 브랜드 이미지 파악 연구 또한 드물다.

브랜드 파워 모형을 기반으로 하는 평가 방법

〈Table 2〉 Anholt's 6 city brands aspect

Aspect	Meaning
Presence	A city's international status and standing. How familiar they are with the city, whether the city made an important contribution to the world in culture, science.
Place	Physical aspects of a city. How pleasant to travel around the city, how beautiful the city is.
Potential	Economic and educational opportunity of a city. How easy to find a job and do a business, how good for oneself or family member to get a higher educational qualification.
Pulse	The appeal of a vibrant urban lifestyle of a city. How exciting and how interesting to live in or to travel.
People	How warm and friendly of city people to outsiders, How easy to find and fit into a community which shares his own language and culture, How safe to live in.
Prerequisites	Basic qualities of a city. How easy to find satisfactory and affordable accommodation, amenities(e.g. schools, hospitals, public transport, sports facilities)

론 중 가장 보편적으로 사용되는 도시 브랜드 이미지 평가 방법은 안홀트의 방법이다. 안홀트는 세계 도시 브랜드 지수를 측정 발표하면서, 도시 브랜드 이미지를 결정하는 6가지 요소를 다음과 같이 제시하였다. 이 6가지 요소에 대하여 도시 간 비교를 함으로써 각 도시의 브랜드 가치가 어느 정도인지가 평가된다.

그 동안 장소 브랜드 이미지 분석에 있어 기존의 방법은 설문 조사, 면담, 전문가 조언 등의 방법으로 이루어져왔다. 하지만 이는 적시적이거나 평가를 희망하는 시점에서의 기민한 분석이 어렵고, 사람에 의한 평가를 진행하므로 조사에 있어서 고비용 구조이며, 면담과 전문가 조언이 주관적일 가능성이 있어 비용 효과적으로 적시에 도시의 이미지를 분석하기에는 한계가 있다. 따라서 이에 대한 개선 방향으로 텍스트마이닝을 고려할 수 있는데, 이의 장점으로는 짧은 시간 안에 분석이 가능하고 객관적인 데이터를 사용하며, 현장 경험이 있는 일반인의 견해를 반영할 수 있고 또한 비용이 적다는 것 등이 있다.

3. 방법론

3.1. 이론적 배경

본 연구에서 제안하는 방법론은 브랜드 파워 모형에 이론적 기반을 둔 안홀트의 방법론을 배경으로 하고 있다. 하지만 안홀트 방법론과의 차별점은 첫째, 브랜드 인식(brand awareness)나 브랜드 이미지의 차원 및 브랜드 자산을 측정하는 항목에 대한 측정을 빅데이터에 대한 텍스트마이닝 기법으로 수행한다는 것이다. 둘째, 평가 항목을 설문항목 그대로 사용하지 않고 안홀트

의 평가 차원에서 등장하는 문장 내의 키워드들로 대체하였다.

분석 과정은 다음과 같다. 첫 번째, 이미지를 탐색할 장소를 선정한다. 본 논문에서는 도시 수준으로 하였다. 두 번째, 장소 브랜드의 가치를 결정하는 6가지 요소와 관련된 유사 단어를 선정한다. 세 번째, 장소 관련 웹 페이지 콘텐츠를 검색 사이트를 활용하여 자동으로 수집한다. 이를 위해 먼저 탐색할 장소와 결정 요소의 유사 단어를 조인(join)하여 쿼리를 만들고 검색 사이트(본 논문에서는 Google.com)의 검색 결과를 프로그래밍적으로 얻기 위하여 Custom Search API 계정을 만든다. 다음으로 쿼리를 Custom Search API에 넣어 검색 결과를 JSON 타입으로 확보하고 JSON을 분석하여 해당 장소에 대한 구글의 검색 결과 URL을 확보한다. 이렇게 획득한 각 URL 주소로 접근하여 웹 페이지 내용을 다운로드 받는다. 네 번째로 수집된 비정형 자료에 대한 전처리를 한다. 전처리는 저장한 웹 페이지 내용에서 HTML 및 Link와 관련된 정보를 제거하고, 저장된 내용을 문장부호(마침표, 물음표, 느낌표)와 다양한 기호(○, ►, □, ▲ 등)로 문장 분리하며, 부정 극성에 의한 오류를 방지하기 위하여 부정소(못, 많, 아니)가 들어간 문장을 제외한다. 또한 노이즈를 줄이기 위하여 해당 장소(예: 서울) 또는 장소 관련 어휘(예: 장소, 도시, 시내, 타운, 시티, 곳)가 있는 문장(유효 콘텐츠)만을 수집하는 것이 전처리 작업이다. 다섯 번째로 본격적인 브랜드 이미지 평가를 수행한다. 이를 위해 먼저 안홀트의 도시 평가 설문조사 방법을 분석하여 브랜드 이미지의 결정요소 별 유사 단어를 선정한 후, 형태소 분석기로 유효 콘텐츠에서 전체 어절 수, 불용어 빈도, 결정요소 별 유사 단어 빈도를 계산한다. 이를 통해 결정요소 별 유사단

어의 빈도를 전체 어절 수에서 불용어 빈도를 빼 값으로 나누어 결정요소 별 브랜드 이미지의 인덱스를 구할 수 있다. 이와 같은 단계가 마무리되면 최종적으로 비교 대상 도시와 우열을 비교한다.

3.2. 유사 단어 선정

브랜드 이미지와 관련된 유사 단어는 브랜드

인덱스 관련하여 가장 널리 알려진 안홀트의 전략 요소에 대한 평가 항목에 대한 정의서에 표현된 문구에서 선정하였다. 문구에서 직접 번역이 되는 단어들(Direct Words)과 간접 번역이 되는 단어들(Indirect Words)를 추출하여 한국어로 작성된 문서에서 브랜드 이미지를 추출할 단어들로 삼았다. 그 결과는 <Table 3>과 같다.

<Table 3> Anholt's city brand aspects' contents and its similar words

Aspects	Classification	Meaning / Words
The Presence	Meaning	This point of the city brand hexagon is all about a city's <u>international status</u> and standing. People are asked how <u>familiar</u> they are with each of the 30 cities in the survey, whether they have actually visited them or not and what the cities are <u>famous</u> for. They are also asked whether each city has made an important <u>contribution</u> to the world in <u>culture, science</u> or in the way cities are governed during the last 30 years.
	Direct Words	Kookje(국제), Segye(세계), Myungsung(명성), Youmyung(유명), Iksuk(익숙), Chinmil(친밀), Chinsuk(친숙), Chingun(친근), Kiyeo(기여), Kongheon(공헌), Moonhwa(문화), Kwahak(과학)
	Indirect Words	Tudureoji(두드러지), Chinmil(친밀), Chinsuk(친숙), Chingeon(친근), Injido(인지도), Yesul(예술), Eumak(음악), Misul(미술), Misulkwan(미술관), Munhak(문학), Yeonguk(연극), Younghwa(영화), Younghwagwan(영화관), Geokjang(극장), Moonhwajae(문화재), Moolihak(물리학), Saengmoolhak(생물학), Chumdan(첨단)
The Place	Meaning	This section explores people's perceptions about the <u>physical aspects</u> of each city: how <u>pleasant</u> or unpleasant they <u>imagine</u> it is to be outdoors and to <u>travel</u> around the city, how <u>beautiful</u> it is and what the climate is like.
	Direct Words	Oegwan(외관), Oehyung(외형), Yeohaeng(여행), Sunho(선호), Image(이미지), Migwan(미관), Arumdap(아름답), Youkwae(유패), Kwaejuk(쾌적), Mark(맑)
	Indirect Words	Junbyeong(전경), Kyeongchi(경치), Myeongso(명소), Yakyoung(야경), Yeopp(예쁘), Mutji(멋지), Jo(좋), Joha(좋아하)
The Potential	Meaning	This point of the city brand hexagon considers the <u>economic</u> and <u>educational opportunities</u> that each city is believed to offer visitors, business and immigrants. The panelists are asked how easy they think it would be to <u>find a job</u> in the city, and, if they had a business, how good a place they think it would be to do <u>business</u> in. Finally, they are asked whether each city would be a good place for them or other family members to get a <u>higher educational</u> qualification.
	Direct Words	Business(비즈니스), Kyoyuk(교육), Kihwae(기회), Kanung(가능), Jamjaeryuk(잠재력), Kyungje(경제), Kujik(구직), Jikup(직업), Jikjang(직장), Iljari(일자리), Taehak(대학), Taehakyo(대학교), Taehakwon(대학원), Saup(사업)

Aspects	Classification	Meaning / Words
The Potential	Indirect Words	Kyoyukyeol(교육열), Chuiup(취업), Taehakga(대학가), Taehakro(대학로), Taehakin(대학인), Taehaksaeng(대학생), Hakbu(학부), Saupga(사업가), Tuja(투자), Muyeok(무역)
The Pulse	Meaning	The appeal of a <u>vibrant urban</u> lifestyle is an important part of each city's brand image. This section explores how <u>exciting</u> people think the cities are, and how easy they think it would be to find <u>interesting</u> things to do, both as a short-term visitor and as a long-term resident.
	Direct Words	Tosijuk(도시적), Tohoeji(도회지), Tosihyeong(도시형), Julgup(즐겁), Julgeoum(즐거움), Jaemi(재미), Jaemiit(재미있), Jaemit(재밌), Hwalbal(활발), Heungbun(흥분), Heungmi(흥미), Heungmirop(흥미롭), Heungmijinjin(흥미진진), Saengdong(생동), Saengdonggam(생동감), Yeokdong(역동)
	Indirect Words	Maeryuk(매력), Hwalgi(활기), Hwalgicha(활기차), Hwalryuk(활력), Hwalryukso(활력소)
The People	Meaning	The people make the city, and in this point of the hexagon respondents are asked if they think the inhabitants would be warm and <u>friendly</u> , or cold and prejudiced against outsiders. They are asked whether they think it would be easy for them to find and <u>fit into a community</u> which shares their language and culture. Finally, and very importantly, panelists are asked how <u>safe</u> they think they would feel in the city.
	Direct Words	Chinjul(친절), Anjun(안전), Chian(치안), Eoulri(어울리)
	Indirect Words	Chakha(착하), Wooho(우호), Gongpyung(공평), Gaebang(개방), Sotong(소통), Tonghwa(동화), Jukeung(적응)
The Prerequisites	Meaning	This sections asks people about how they perceive the basic qualities of the city: what they think it would be like to live there, how easy they think it would be to find <u>satisfactory, affordable accommodation</u> and what they believe the general standard of public amenities is like – <u>schools, hospitals, public transport, sports facilities</u> and so on.
	Direct Words	Konggong(공공), Taejung(대중), Kicho(기초), Kiban(기반), Manjok(만족), Heupjok(흡족), Almat(알맞), Jukdang(적당), Jukjul(적절), Jugeo(주거), Geoju(거주), Geojuji(거주지), Jip(집), Jutaek(주택), Gajungjip(가정집), Hakgyo(학교), Byungwon(병원), Uiryoo(의료), Kyotong(교통), Sports(스포츠), Kyunggijang(경기장)
	Indirect Words	Kodeunghakgyo(고등학교), Junghakgyo(중학교), Chodeunghakgyo(초등학교), Youchwon(유치원), Eorinijip(어린이집), Bus(버스), Jihachul(지하철), Taxi(택시), Kongwon(공원)

3.2. 장소 관련 웹 페이지 콘텐츠 수집

위의 과정 중 Google API를 이용하여 장소의

검색 결과 웹 페이지를 받는 주요 코드는 <Figure 2>와 같다.

```

## Program codes for text analysis of specific place via Google search engine.
//Get result by JSON type when keywords query is sent to Google search engine.
String query = googleURL+apiKey+searchKey+keyword+json+pageNum+resultNum;
URL url=new URL(query); ## Sending a query
URLConnection conn=(URLConnection)url.openConnection();
conn.setRequestMethod("GET");
BufferedReader br=new BufferedReader(new InputStreamReader(conn.getInputStream()));
StringBuffer buffer=new StringBuffer();

while((line=br.readLine())!=null)
buffer.append(line); ## Getting a result by JSON type

//Extract place website's addresses by reading JSON file
JSONArray jsonArr = obj.getJSONArray("items");
String[] link= new String[jsonArr.length()]; ## Creating address array

for(int i=0; i< jsonArr.length(); i++)
link[i] = jsonArr.getJSONObject(i).getString("link"); ## Extracting address

//Remove HTML tags in webpages by Jsoup library
for(int i=0; i<link[i].length(); i++)
{
InputStream input = new URL(link[i]).openStream();
Document doc = Jsoup.parse(input, encoding, link[i]); ## Removing HTML tags
text = doc.toString(); ## Saving result
}

//Calculate words frequency by RHINO library
RHINO m = new RHINO();
String output = m.ExternRhinoCall(text, true); ## Analyzing and calculating a words frequency
saveResult(output, "analyseReport.txt", "euc-kr"); ## Saving a result

//Create a frequency graph with R engine and wordcloud library
Rengine R = new Rengine(new String[] {"args"}, false, null); ## Using R engine
R.eval("png(filename=\""+rpath+"analyseReport.png\")");
R.eval("rev <- read.table(\"analyseReport.txt\")");
R.eval("wordcount <- table(rev)");
R.eval("library(\"RColorBrewer\", lib.loc =\""+rLibPath+"\")");
R.eval("palette <- brewer.pal(9, \"Set1\")");
R.eval("library(\"wordcloud\", lib.loc =\""+rLibPath+"\")");
R.eval("wordcloud(names(wordcount), freq=wordcount, scale=c(5,0.5), rot.per=0.25, min.freq="+2+",
random.order=F, random.color=T, colors=palette)"); ## wordcloud 분석
R.eval("dev.off()"); ## Exiting a program and saving a result

```

〈Figure 2〉 Main method for collecting atypical data

빈도 계산 시에는 불용어(stopwords)는 포함시키지 않는다. 본 연구에서 불용어는 다음과 같이 정의하였다. 첫째, 형식어휘. 대명사, 의존명사, 어미, 접속사 등 실질적인 의미를 가지고 있지 않은 어휘들은 제거하였다. 둘째, 말뭉치 상의 고빈도 어휘. 실질어휘라고 하더라도 ‘하다, 있다, 되다’와 같이 말뭉치 상의 고빈도 어휘는 실제적인 의미는 많이 갖고 있지 않으면서 많은 문장에서 발견되므로 제거한다. 본 연구에서는 국립국어원 1,200만 어절 말뭉치의 어휘빈도 40위까지 중 실질어휘(사람, 때, 년, 말, 일, 하, 있, 되, 지, 보, 같, 대하, 주, 위하, 때문, 말하, 가, 받, 그렇, 알)를 대상으로 하였다 (Kang and Kim, 2009). 셋째, 지역별 분석 결과에서 공통적으로 나타나는 고빈도 어휘. 예를 들어, ‘함평’과 관련된 문서에서는 ‘함평’, ‘전남’, ‘전라남도’가 반복적으로 나타나며 이들 단어로는 지역의 특색을 살펴보기 어려우므로 제거한다.

3.4. 브랜드 이미지 평가

특정 도시에 대한 브랜드 이미지 평가를 위해서 다음과 같은 순서로 실행한다.

Step 1: 검색 포털(예: 구글 검색)에 특정 도시(i)에 대해서 도시명(c_i)과 특정 평가 항목(e_{jk})으로 AND 검색을 하여 n개의 웹 페이지를 다운로드 한다. 이 n개의 웹 페이지로부터 특정 도시(i)에 대해서 특정 평가 차원(j)의 k번째 항목의 등장 횟수인 f_{ijk} 를 구한다. 즉,

$$f_{ijk} = \sigma_{c_i \otimes e_{jk}}$$

이때 $\sigma_{c_i \otimes e_{jk}}$ 는 c_i 과 e_{jk} 에 대해서 AND 검색 (\otimes)하여 도출한 빈도수이다.

이렇게 하여 모든 c_i 에 대해 그리고 모든 평가 차원(j)과 그 차원의 모든 항목(k)에 대해 f_{ijk} 를 구한다.

Step 2: 각 평가 항목에 대해서 정규화를 실시한다. 정규화된 빈도를 의미하는 \tilde{f}_{ijk} 는,

$$\tilde{f}_{ijk} = \frac{f_{ijk} - \min_k(f_{ijk})}{\max_k(f_{ijk}) - \min_k(f_{ijk})}$$

로 구한다.

이렇게 정규화 하는 이유는 평가 항목별로 자주 등장하고 그 편차가 큰 단어와 그렇지 않은 단어가 존재하기 마련이며 이러한 편차 및 규모로부터 유발되는 편의(bias)를 제거하기 위함이다.

Step 3: 특정 평가 차원(j)에 대해서 도시의 브랜드 이미지 크기(\tilde{f}_{ij})를 다음과 같이 계산한다.

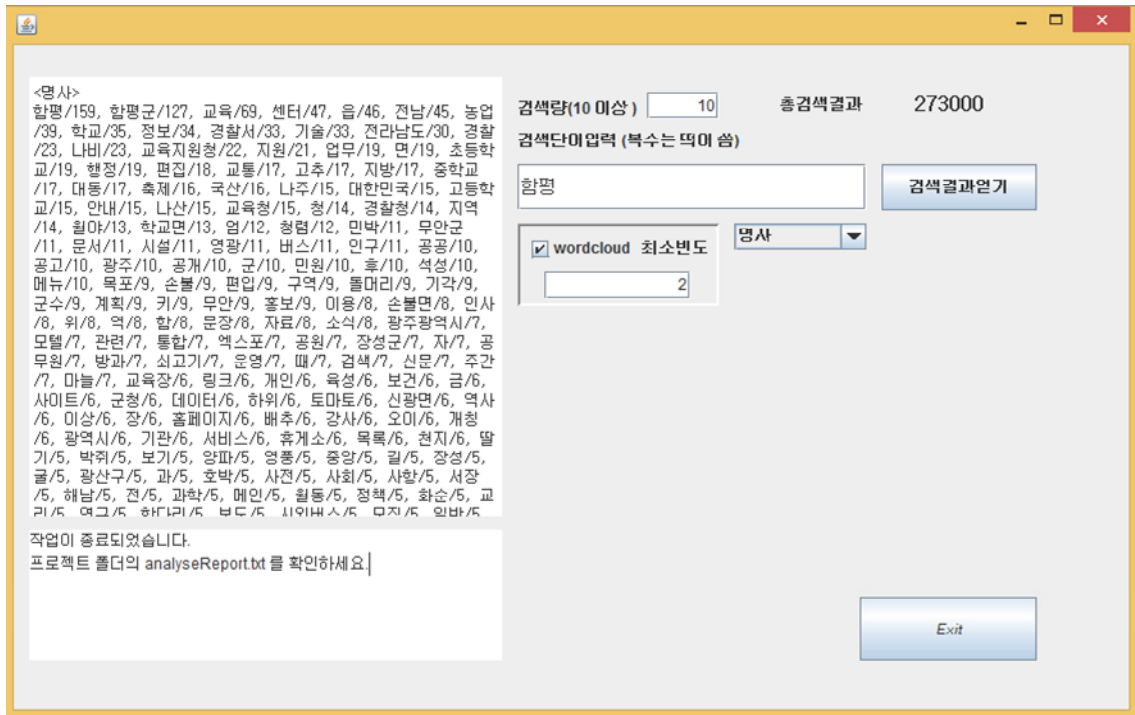
$$\tilde{f}_{ij} = \sum_{\forall k} w_{jk} \tilde{f}_{ijk}$$

단 여기서 w_{jk} 는 특정 평가 항목이 차지하는 상대적 중요도($0 \leq w_{jk} \leq 1$, $\sum_k w_{jk} = 1$)이다.

4. 실험

4.1. 구현

본 논문의 방법론의 실현 가능성을 보이기 위



<Figure 3> Result of morphological frequency of place in text format

해 장소 브랜드 이미지의 구축을 위해 비정형 자료로부터 이미지 추출을 하는 시스템을 개발하였다. 시스템은 Java application으로 되어 있으며 특정 장소(지자체, 관광지, 건물 등)를 입력하고 검색 개수를 지정하면 Google API를 활용하여 구글 사이트로부터 장소 관련 웹 문서를 받아오게 된다. 수집된 웹 문서는 자체 개발한 형태소 분석기 RHINO에 의하여 형태소 빈도계산이 되고, 그 결과는 다시 시스템으로 전달되어 무의미 단어(stopping words) 및 HTML 태그와 태그 안의 링크와 관련된 내용을 삭제한다. 그 후에 각 브랜드 이미지 평가 항목 별로 등장횟수(frequency)를 계산하여 브랜드 이미지 강도를 구한다.

<Figure 3>은 구현된 시스템을 활용하여 형태소의 등장횟수를 정렬해 보인 예이며, <Figure 4>



<Figure 4> Result of morphological frequency of place in wordcloud

는 이를 워드클라우드로 변형하여 보인 예이다.

4.2 실험방법

본 논문에서 제안한 방법의 성능을 보이기 위해 가장 널리 알려진 평가 방법인 안홀트 방법을 포함하여 다음과 같은 네 가지 방법과 비교하였다.

방법1: 임의로 순위를 매기는 방법 (RANDOM).

방법2: 안홀트의 설문방식대로 일반인이 평가하는 방법 (HUMAN). 이를 위해 도시에 대한 이미지를 평가할 수 있는 인문사회계열의 석박사 과정생 9명을 대상으로 안홀트의 도시 평가 방법론 및 각 평가 차원에 대해서 충분히 설명을 한 후에 설문지 방식으로 각 도시 별 여섯 가지 평가 차원에 대해서 7점 척도(1: 매우 낮음, 7: 매우 높음)를 기입하도록 하였다. 그런 후에 9명의 평가 결과의 평균치를 가지고 각 도시의 순위를 도출하는 방법이다.

방법3: 본 논문의 방법을 사용하며 모든 평가 항목을 다 반영하는 방법 (TM). 도시에 관한 문서에 안홀트(2006)의 각 평가 차원에 대한 정의서에 등장하는 단어들이 어느 정도의 비율로 등장하는지를 계산한다. 먼저 구글 검색 엔진을 이

용하여 도시에 대한 문서를 추출한 뒤, 3.1에 설명된 방법과 같이 부정소가 들어간 문장은 제외하고, 장소 관련 어휘가 들어간 문장만 추출하여 유효 콘텐츠를 확보한다. 다음으로, 3.2에서 정리한 안홀트의 도시 평가 설문에 사용된 결정요소별 유사단어의 빈도를 전체 어절 수에서 불용어의 수를 제외한 값으로 나누어 결정요소별 브랜드 이미지의 인덱스를 얻는다. 그리고 최종적으로 각 평가 차원 별로 도시 이미지 점수를 정렬하여 순위를 도출하는 방법이다.

방법4: 본 논문의 방법을 사용하되 안홀트의 방법으로 학습한 것으로 유의한 것으로 추정되는 평가 항목만을 반영하는 방법 (TM-LEARN). 방법3(TM)과 같이 각 항목을 선정하되 학습 셋에서 안홀트의 순위와 상관관계가 낮은 것을 제외한 일부 항목들만을 가지고 테스트하는 방법을 사용하였다.

평가 측정은 RMSE (Root Mean Square Error)를 사용하였다. 즉, 안홀트에 의하여 발표된 도시 브랜드 이미지 순위와 각 방법에서 제안하는 순위 사이의 RMSE를 구하여, 낮은 값을 가질수록 우수한 측정 방법이 되는 것이다. 평가에 사용된 도시는 서울, 홍콩, 뉴욕 등 안홀트 평가에 들어있는 50개 모두이다.

〈Table 4〉 Result of selection of statistically meaningful evaluation elements in TM-LEARN experiment

Aspect	Meaningful evaluation elements	Dropped out elements
Presence	World, Famous, Familiar, Culture, Art, Museum	Intimate, Reputation, Contribution
Place	Travel, Landscape, Gorgeous, Good, Pleasing, Comfortable	Image, Sightseeing, Beauty, Pretty, Prefer
Prerequisite	Base, Satisfaction, Suitable, Relevant, House	Public, Infrastructure, Dwelling, School, Arena
People	Kindness, Communication, Secure, Assimilation, Harmony	Favor, Open, Fair, Adaptation
Pulse	Civilized, Fun, Interesting, Vivid, Dynamic	Attractive, Joyful, Exciting
Potential	Education, Recruit, Possibility, Student, Investment	Business, Schooling, Opportunity, Potential, University

4.3. 결과

위의 실험 방법에 근거하여 네 가지 대안에 대해서 실험을 수행하였다. 특히 학습 데이터를 통해서 TM-LEARN을 수행한 결과 여섯 차원에 대해서 안홀트에 의한 평가에 유의하게 사용될 수 있는 평가 항목들은 <Table 4>와 같이 도출하였다. 이 항목들은 안홀트에 의한 평가 결과인 도시 브랜드 평가 순위와 통계적으로 유의하게 정비례 또는 반비례 하는 항목들이다.

안홀트 방법론에서 제시한 6가지 평가 차원에 대해서 <Table 5>와 같은 결과가 나왔다. 그 결과 제안한 방법(TM-LEARN)이 RANDOM, HUMAN 방법보다 RMSE 측면에서 우수한 결과를 보였다. 한편 TM 방법은 HUMAN 방법과 별다른 차이가 없었다. 따라서 본 연구에서 제안한 방법론인 TM-LEARN은 일반인에 의한 평가보다 더 정확할뿐더러 비용 효과적인 방법이라고 결론을 내릴 수 있다.

5. 토의 및 결론

본 논문은 공개된 비정형자료를 통해 장소 브

랜드 이미지 및 그 강도를 자동적으로 획득하는 방법론을 제안하고 실제 평가 결과와 비교 검증하였다. 제안된 방법론은 첫째, 일반인들을 대상으로 하는 평가 방법에 비하여 더욱 정확한 것으로 나타났다. 둘째, 자동화 방법을 사용했기 때문에 현재 안홀트가 진행하고 있는 방식인 전통적인 설문 조사 방식으로 많은 사람이 동원되고 많은 비용이 수반이 되는 것과 비교해볼 때, 제안된 방법론이 평가에 소요되는 시간 및 비용이 더 효율적이면서도 안홀트 방법에 의하여 평가된 도시 브랜드 이미지 강도를 상대적으로 정확하게 추론하였다. 셋째, 제안 방법론은 수시로 평가가 가능하기 때문에 연간으로 강도를 측정하는 안홀트 방법에 비하여 더욱 기민 또는 적시적이다. 이는 장소 브랜딩 과업이 지속적으로 이루어져야 의미 있음을 고려해볼 때 중요한 장점이다. 다음으로 안홀트 등의 방법은 이미 지정된 도시들에 대해서만 평가를 하는 것에 비하여 제안 방법론은 어떤 도시나 다른 장소에 대해서도 그대로 적용 가능하다. 마지막으로, 설문 방식이 주관적이며 응답자의 특성에 따라 다른 결과가 나올 가능성이 높는데 비하여 제안 방법은 공개된 자료를 바탕으로 하기 때문에 상대적으로 객

<Table 5> Comparison result of RMSE performance (Lower is better)

Aspect	RANDOM	HUMAN	TM	TM-LEARN
Presence	21.6	19.4	19.9	18.2
Place	19.9	18.6	15.2	16.9
Prerequisite	21.3	18.8	19.8	19.6
People	21.3	19.3	22.4	16.9
Pulse	21.0	18.1	20.9	17.6
Potential	21.5	21.2	21.2	16.6
Mean (std)	21.1 (0.6)	19.2 (1.1)	19.9 (2.5)	17.6 (1.1)

관성을 더 유지할 수 있다. 결국, 도시 이미지 평가에도 텍스트마이닝 접근법이 정확성, 비용효율성, 적시성, 확장성, 그리고 신뢰성 측면에서 유효함을 보일 수 있었다. 따라서 본 연구에서 제안한 방법론은 안홀트 방식에 상호 보완적으로 사용될 수 있을 것이다.

본 논문의 방법은 실무적으로 의미가 있다. 지방 자치단체의 입장에서 자신의 지역의 전략 방향을 정하고 장소의 브랜드 이미지를 높이는 작업을 할 때 이에 대한 측정 방법으로서 설문 조사 등의 전통적인 방법을 진행하고, 이에 보완적으로 본 논문의 방법론을 적용하여 설문조사에서 간과할 부분을 보강해줄 수 있는 좋은 대안이 될 것이다. 따라서 빅데이터의 시대를 맞이하여 기존의 전통적인 조사 방식과 새로운 빅데이터의 접목이라는 측면에서 장소 브랜드 이미지를 측정하는 데 있어서 상호 보완적인 좋은 사례가 될 것이다. 둘째로, 제안 방법론을 기반으로 하여 형태소 분석, 정형화, 장소 브랜드의 속성 (facet) 추출 및 강도나 빈도분석을 할 수 있으므로, 이 방법론을 통해 추론되는 이미지 결과는 장소 마케팅 전략을 수립하거나 도시공학 전문가들에게 직관을 제공하여 보다 더 풍부하거나 쉽게 전문가 작업을 할 수 있도록 도울 수 있다.

향후에는 장소 브랜드 이미지를 형성하는 속성 별로 등장횟수를 계산 한 후에 장소 브랜드에 대한 태도, 연상, 그리고 브랜드 자산과의 인과 관계를 자동으로 파악할 수 있는 부분까지 구현하고 실증적 실험을 할 예정이다.

참고문헌(References)

Aaker, D. A. and K. L. Keller, "Consumer Evaluations

of Brand Extensions," *The Journal of Marketing*, Vol.51, No.1(1990), 27~41.

Aaker, D. A., *Building strong brands*, FreePress, NY, 1995.

Anholt, S., "The Anholt-GMI City Brands Index: How the world sees the world's cities," *Place Branding*, Vol.2, No.1(2006), 18~31.

American Marketing Association, *Marketing Definitions: A Glossary of Marketing Terms*, American Marketing Association, Chicago, 1960.

Chang, D.R., M.Y. Jun and S.K. Kwon, "A Study on Place Branding for Promotion of the Value of City Brand," *Journal of Korean Society of Design Science*, Vol.23, No.4 (2010), 351~366.

Cuneo, Alice A., "Sears To Sell Goods On Hispanic Web Site: Looks To Build On Brand Equity In Latin America," *Crain's Chicago Business* (1999, Oct 36).

Faircloth, J. B., L.M. Capella, and B.L. Alford, "The Effect of Brand Attitude and Brand Image on Brand Equity," *Journal of Marketing Theory and Practice*, Vol.9, No.3(2001), 61~75.

Gu, J. and J. Lee, "A Study on the Evaluation and Indexation of City Brand Equity," *Productivity Review*, Vol.22, No.3(2008), 179~181.

Hankinson, G., "Relational network brands: Towards a conceptual model of place brands," *Journal of vacation marketing*, Vol.10, No.2(2004), 109~121.

Kang, B. and H. Kim, *Use-Frequency of Korean Language*, Hankookmunhwasa, 2009.

Keller, K. L., "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity," *The Journal of Marketing*, Vol.57, No.1(1993),

- 1~22.
- Kim, H.-j., N.-o. Jo, and K.-s. Shin, "Text Mining-Based Emerging Trend Analysis for the Aviation Industry," *Journal of Intelligence and Information Systems*, Vol.21, No.1(2015), 65~82.
- Lee, J. H. and K. H. Lim, "A Case Analysis of Place Branding Using Public Cultural Contents," *Journal of Basic Design & Art*, Vol.12, No.6(2011), 317~325.
- Lee, K. H., "A Study on the Features of Place Brand Imaging for Creating Shared Value," *Journal of Basic Design & Art*, Vol.15, No.5(2014), 423~431.
- Lee, T. H., "Brand index Evaluation of Tourism brand near the Capital area," *International Journal of Tourism and Hospitality Research*, The Korea Tourism Research Association, Vol.19, No.3(2005), 471~484.
- Lee, W. and N. Lim, "A Study on the Elements of City Brand Image and Influences," *Journal of Korea Planners Association*, Vol.40, No.6(2005), 177~192.
- Na, W., and R. Marshall, and K. L. Keller, "Measuring brand power: validating a model for optimizing brand equity," *Journal of product & brand management*, Vol.8, No.3(1999), 170~184.
- Na, W., and R. Marshall, "Brand power revisited: measuring brand equity in cyber-space," *Journal of product & brand management*, Vol.14, No.1(2005), 49~56.
- Newman, J. W., "New Insight, New Progress, For Marketing," *Harvard Business Review*, November-December(1957), 95~102
- Park, S. H. and D. R. Chang, *Place branding: from Hongdae to London*, Design House, 2009.
- Yi, J. B., C. K. Lee, and K. J. Cha, "An Analysis of IT Trends Using Tweet Data", *Journal of Intelligence and Information Systems*, Vol.21, No.1(2015), 143~160.

Abstract

An Efficient Estimation of Place Brand Image Power Based on Text Mining Technology

Sukjae Choi* · Jongshik Jeon* · Biswas Subrata** · Ohbyung Kwon***

Location branding is a very important income making activity, by giving special meanings to a specific location while producing identity and communal value which are based around the understanding of a place's location branding concept methodology. Many other areas, such as marketing, architecture, and city construction, exert an influence creating an impressive brand image. A place brand which shows great recognition to both native people of S. Korea and foreigners creates significant economic effects. There has been research on creating a strategically and detailed place brand image, and the representative research has been carried out by Anholt who surveyed two million people from 50 different countries. However, the investigation, including survey research, required a great deal of effort from the workforce and required significant expense. As a result, there is a need to make more affordable, objective and effective research methods. The purpose of this paper is to find a way to measure the intensity of the image of the brand objective and at a low cost through text mining purposes. The proposed method extracts the keyword and the factors constructing the location brand image from the related web documents. In this way, we can measure the brand image intensity of the specific location. The performance of the proposed methodology was verified through comparison with Anholt's 50 city image consistency index ranking around the world. Four methods are applied to the test. First, RNADOM method artificially ranks the cities included in the experiment. HUMAN method firstly makes a questionnaire and selects 9 volunteers who are well acquainted with brand management and at the same time cities to evaluate. Then they are requested to rank the cities and compared with the Anholt's evaluation results. TM method applies the proposed method to evaluate the cities with all evaluation criteria. TM-LEARN, which is the extended method of TM, selects significant evaluation items from the items in every criterion. Then the method evaluates the cities with all selected evaluation criteria. RMSE is used to as a metric to compare the evaluation results. Experimental

* Humanitas BigData Research Center, KyungHee University

** School of Management, KyungHee University

*** Corresponding author: Ohbyung Kwon

School of Management, KyungHee University

26 Kyunghee-daero, Dongdaemun-gu, Seoul 130-701, Korea

Tel: +82-2-961-2148, Fax: +82-2-961-2148, E-mail: obkwon@khu.ac.kr

results suggested by this paper's methodology are as follows: Firstly, compared to the evaluation method that targets ordinary people, this method appeared to be more accurate. Secondly, compared to the traditional survey method, the time and the cost are much less because in this research we used automated means. Thirdly, this proposed methodology is very timely because it can be evaluated from time to time. Fourthly, compared to Anholt's method which evaluated only for an already specified city, this proposed methodology is applicable to any location. Finally, this proposed methodology has a relatively high objectivity because our research was conducted based on open source data. As a result, our city image evaluation text mining approach has found validity in terms of accuracy, cost-effectiveness, timeliness, scalability, and reliability.

The proposed method provides managers with clear guidelines regarding brand management in public and private sectors. As public sectors such as local officers, the proposed method could be used to formulate strategies and enhance the image of their places in an efficient manner. Rather than conducting heavy questionnaires, the local officers could monitor the current place image very shortly a priori, than may make decisions to go over the formal place image test only if the evaluation results from the proposed method are not ordinary no matter what the results indicate opportunity or threat to the place. Moreover, with co-using the morphological analysis, extracting meaningful facets of place brand from text, sentiment analysis and more with the proposed method, marketing strategy planners or civil engineering professionals may obtain deeper and more abundant insights for better place brand images. In the future, a prototype system will be implemented to show the feasibility of the idea proposed in this paper.

Key Words : Text Mining, Brand Image, Location Brand, Keyword Analysis, Anholt's Brand Index

Received : June 1, 2015 Revised : June 15, 2015 Accepted : June 16, 2015

Type of Submission : Fast Track Corresponding Author : Ohbyung Kwon

저 자 소개



최석재

현재 경희대학교 후마니타스 빅데이터 연구센터 전임연구원으로 재직 중이다. 고려대학교 국어국문학과에서 국어학으로 학사, 석사, 박사 학위를 취득하였고, 미국 Carnegie Mellon Univ. Language Technologies Institute 객원연구원, 중국 연변과학기술대학 언어공학연구소 실장, 고려대학교 BK21 연구교수, 성신여자대학교 국어국문학과 초빙교수 등을 거쳤다. 한국어 정보를 전산화하는 방안과 비정형 빅데이터의 분석에 관심이 있다.



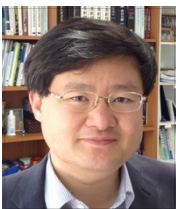
전종식

현재 경희대학교 경영대학 석박사 통합과정으로 재학 중이다. 서울대학교 경영학과에서 학사(1988)를 취득했다. 최근 관심분야는 빅데이터 분석과 공공정책 등이다.



비스워스 수브르타

경희대학교 일반대학원 경영학과 박사과정 재학중이다. 상명대학교 경영학과 학사(2010) 및 정보보안 분야로 석사(2014)학위를 취득하였다. 주요논문으로는 AHP 기법을 활용한 Big Data 정보보안관리 요소들의 우선순위 분석에 관한 연구 등이 있다. 관심 연구분야는 경영정보, 빅데이터 분석, 정보보안관리체계, 경영정보보호, 텍스트마이닝 등이다.



권오병

현재 경희대학교 경영대학에서 교수로 재직 중이다 (전공 MIS). 서울대학교 경영학과에서 경영학사 (1988), 한국과학기술원 경영과학과에서 각각 공학석사(1990)와 공학박사 (1995)를 취득하였다. 1995년에는 중국연변과기대 경영정보학과 조교수로 재직한 바 있으며, 1996년~2004년 동안 한동대학교 경영경제학부 부교수로도 재직한 바 있다. 그 연구는 Journal of Management Information System, International Journal of Information Management, Decision Support System, 지능정보연구 등에 다수 발표되었다. 현재 관심분야로는 빅데이터 분석, IoT, 유비쿼터스 컴퓨팅, ICT 융합 등이 있다.