

사회연결망서비스의 EXIF 기반 Hashtag 추천 시스템

EXIF-based Hashtag Recommender System on Social Networking Service

이 상 훈 (Sang Hoon Lee) 대구대학교 정보통신대학 컴퓨터정보공학부
김 수 연 (Su-Yeon Kim) 대구대학교 정보통신대학 컴퓨터정보공학부, 교신저자

요 약

많은 사용자들이 자신의 일상과 활동을 사회연결망서비스(SNS)에 업로드하고 있으며 자신들의 포스팅을 설명하기 위한 Hashtag를 사용하고 있다. Hashtag는 사용자가 스스로 자신의 포스팅에 대한 카테고리를 지정한다는 장점이 있으나 최근까지도 사용자가 직접 수동적으로 입력해야 한다는 번거로움이 있었다. 이에 본 연구에서는 이러한 문제를 개선하기 위하여 SNS 상에서 사용자가 업로드 하는 이미지를 기반으로 하여 적절한 Hashtag를 추천하는 방법을 제안하였다. 본 연구를 위하여 촬영정보에 기반한 분석, 주소에 기반한 분석, 타사용자에 기반한 분석, 이미지 자체에 기반한 분석 방법을 설계 및 구현하였다. 또한 제안 방법이 기존의 시스템에 비해 개선되었는지 확인하기 위하여 성능 테스트를 실시하였으며 15개국 212명의 SNS 사용자를 대상으로 평가를 진행하였다. 분석 결과 기존에 서비스되는 Hashtag 추천 시스템에 비해 제안된 시스템이 높은 정확도의 추천 결과를 보였으며 기존 대비 개선되었다는 것을 확인할 수 있었다.

키워드 : 해시태그, 추천 시스템, EXIF, 사회연결망서비스

I. 서 론

SNS는 사용자가 온라인상에 자신이 직접 작성한 콘텐츠를 업로드 하여 지인이나 유사한 관심사를 가진 다른 사용자들에게 정보를 공유하거나 의견을 전달하는 서비스로 전 세계에서 폭넓게 사용되고 있다. 이러한 변화를 통해 사람들은 다양한 콘텐츠를 매일같이 업로드하게 되고 그 양과 종류 또한 매우 방대해졌다. 방대해진 데이터를 운영자가 제공해주는 카테고리 형태로 분류하는 것은 사

실상 거의 불가능해졌으며 사용자가 스스로 자신의 콘텐츠를 분류할 수 있는 Hashtag 기법이 Web 2.0의 출현과 함께 대다수의 SNS와 온라인 커뮤니티에 사용되고 있다.

Hahstag는 사용자가 스스로 자신의 콘텐츠에 대한 분류를 지정할 수 있게 됨에 따라 과거의 카테고리 형태와는 다르게 하나의 콘텐츠에 복수개의 Hashtag를 설정하여 연관된 정보를 필요로 하는 사람들에게 쉽게 제공해줄 수 있는 기획의 창이 되었다. 또한 Hashtag는 특정 지역에 대한 일반

적인 검색뿐 아니라 현재의 날씨나 이슈, 사고 등에 대한 정보를 빠르게 확인할 수 있는 경로가 되었으며 특정 키워드와 관련된 정보를 쉽게 획득할 수 있는, 기존의 검색을 뛰어넘는 새로운 형태의 정보 획득 방법이 되었다. 하지만 Hashtag는 사용자가 직접 설정하기 때문에 하나씩 직접 입력해야 하는 번거로움이 생겼다. 이에 본 연구는 이러한 Hashtag를 사용자가 적절하게 사용할 수 있도록 키워드를 추천하고 사용자의 선택에 따라 개인화할 수 있는 EXIF(Exchangeable Image File Format) 기반의 Hashtag 추천 방법을 제시하고자 하였다.

본 연구에서는 이미지 내 포함된 EXIF 정보를 추출하고 이를 기반으로 한 추천을 실시한다. EXIF는 이미지에 대한메타데이터를 포함하고 있으며 스마트폰이나 사진기의 GPS 정보도 함께 포함하고 있어 촬영 위치도 확인할 수 있다. 이에 사진의 위도와 경도 정보를 추출하고 이를 이용하여 사용자가 촬영한 곳의 명칭과 지리적 주소 등을 추출한다. 또한 에펠탑 같은 대형 구조물의 경우 촬영자의 위치가 매우 넓은데 이를 파악하기 위해 사용자가 촬영한 인근에서 가장 인기 있는 지역을 파악하고 이를 추천에 이용함으로써 그 정확도를 높인다.

이미지의 위치정보뿐 아니라 타 사용자들이 동일한 위치에서 촬영한 사진을 SNS에 업로드 할 때 사용한 Hashtag를 찾아 추출하여 가장 많이 사용했고 최근에 사용한 Tag를 기반으로 사용자에게 추천할 수 있도록 한다. 또한 이미지 자체를 분석하여 이미지 내 포함된 오브젝트나 배경 등과 관련된 Tag를 추출, 사용자들에게 추천할 수 있도록 한다. 위의 방법으로 추출한 Hashtag는 사용자의 선호도를 고려하여 최적의 추천이 이루어질 수 있도록 개인화된 방식으로 추천하는 시스템을 구현하고자 한다.

II. 이론적 배경

2.1 Hashtag

Hashtag란 Hash(#) 기호에 특정한 문자열을 더

하여 일종의 메타데이터 역할을 하는 Tag를 말한다. 2007년 디자이너인 Chris Messina가 자신의 트위터 계정에 Hash 기호를 이용하여 그룹을 구성하자는 제안 글을 올렸고 트위터 운영자들이 이 기능을 구현하면서 현재의 Hashtag 기능이 시작되었다. 페이스북과 인스타그램, 구글+ 등의 주요 SNS 또한 유사한 기능을 추가하면서 현재와 같이 폭넓게 사용되고 있다(Bennett, 2014).

Hashtag는 # 기호 뒤에 문자열을 붙여 만들며 하나의 게시물 안에 복수 개를 사용할 수 있다. 최근에는 Hashtag 사용자들이 늘어나자 관련 정보를 얻기 위한 검색이나 소통의 경로로 사용되고 있다. 특히 인스타그램이나 핀터레스트 같은 이미지 기반 SNS의 경우 관련 이미지를 쉽고 빠르게 찾을 수 있도록 Hashtag 검색 기능을 지원하고 있으며, Hashtag를 이용하여 현재의 트렌드나 이슈 등을 파악하여 제공하는 서비스도 증가하고 있다(안상욱, 2015). Hashtag 이용자가 늘어나면서 관련 연구도 점차 증가하고 있는 추세이다.

남민지 등(2015)은 인스타그램 내 Hashtag를 이용하여 사용자의 감정을 분류하는 방법에 대해 제안하였으며 김세진, 이상돈(2014)의 연구에서는 SNS 상에서 동시에 출현한 Hashtag 간 연관성을 분석하여 이를 이용한 트렌드 분석 정보를 제공하고자 하였다. 장문종 등(2017)은 해킹된 사용자들 Hashtag 기반으로 찾아낼 수 있는 기술에 대해 연구하였다. 문지훈 등(2015)은 SNS의 Hashtag를 기반으로 하여 목적별 여행지 분류 기법에 대한 연구를 진행하였다. 전민아 등(2013)은 Hashtag를 분석하여 적절한 음악 추천을 위한 방법을 제시하였다. 또한 Moorley and Chinn(2014)의 연구에서는 Hashtag를 이용하여 온라인 커뮤니티를 만들 수 있는 방법을 제시하기도 했다. 이외에도 많은 연구자들이 Hashtag에 대한 연구를 실시하였다(Alarifi et al., 2016; Arnaboldi et al., 2016; Sohn et al., 2015; Yazdanfar and Thomo, 2013).

이처럼 Hashtag에 대한 관심과 활용도가 증가되고 있는 시점에서 본 연구는 사용자들이 일일이

Hashtag를 직접 입력하는 기존의 방법 대신 적절한 Hashtag를 추천해줌으로써 편의성을 높이고 온라인상에서 더욱 많은 메타데이터가 추가되어 또 다른 서비스의 입력 데이터로 활용할 수 있도록 Hashtag 추천 시스템을 설계하고 구현하고자 한다.

2.2 EXIF

EXIF(Exchangeable Image File Format)는 이미지 파일에 포함된 일종의 메타데이터라 할 수 있다. EXIF의 데이터는 이미지의 촬영일시 뿐 아니라 촬영위치, 촬영 시 카메라 설정 등의 정보를 포함하고 있다. EXIF는 오래전부터 사용되고 있던 개념이지만 메타데이터의 개념을 이용하여 디지털 카메라와 스마트폰의 보급이 확산되는 시점에서 기하급수적으로 늘어난 이미지 파일을 처리하기 위한 연구가 많이 진행되었다. 특히 EXIF의 기능을 이용하여 사진을 분류하는 것에 대한 연구가 가장 많이 진행되었다. Liu *et al.*(2005)은 Linear Discriminant Analysis(LDA)를 이용한 분류방법에 EXIF를 적용하여 실내사진과 실외사진 등을 좀 더 정확하게 분류할 수 있는 방법을 연구하였다. Cheng and Shen(2016) 또한 대형 랜드마크 이미지를 분류하는데 EXIF를 하나의 분석요인으로 성능을 높이는데 사용하였다. 류동성 등(2010)과 장철진 등(2006), 제성관 등(2011)의 연구에서도 EXIF의 촬영정보를 디지털 사진에 대한 클러스터링 방법에 사용하였다. 이외에 EXIF를 이용하여 이미지의 저작권을 보호하는 워터마크를 삽입하는 기술 연구도 수행되었다(Huang and Fang, 2010).

EXIF는 GPS 정보를 포함하고 있다. 스마트폰의 경우 촬영한 이미지의 위치정보를 기반으로 사진을 분류하기도 하고 다수의 사용자가 업로드 한 사진으로 전 세계 지도를 만드는 프로젝트 등도 진행되고 있다. 최홍선 등(2012)과 최재용 등(2012)의 연구는 EXIF 내 포함된 GPS 정보를 이용하여 이미지 위치를 표시하고 이를 기반으로 하여 사진을 분류하였다. 이와 같이 위치기반의 서비스들이

점차 증가하고 있다. 본 연구에서는 이미지 위치를 기반으로 하여 사용자에게 적절한 Hashtag를 추천할 수 있는 시스템에 대해 설계하고 분석을 진행한다.

2.3 추천 시스템

추천 시스템이란 수많은 대안 중에서 사용자에게 필요할 것으로 생각되는 정보를 제공해주는 시스템이라 할 수 있다.

일반적으로 추천 시스템은 크게 2가지로 분류할 수 있다. 콘텐츠 자체에 대한 분석을 실시하는 CBF(Content-based filtering)와 유사한 사용자를 기반으로 분석하는 CF(Collaborative filtering)이다.

CBF는 사용자의 프로파일을 기반으로 하여 콘텐츠가 보유하고 있는 속성과 일치하는 정보를 파악하여 사용자에게 추천해주는 기법이다(Pazzani and Billsus, 1997; Shih and Liu, 2008). CBF의 경우 추천하고자 하는 콘텐츠나 서비스 등에 대한 기반 정보와 사용자의 프로파일을 필요로 한다.

CF는 자신과 유사한 사용자 집단이 선호하는 상품이나 서비스 등을 추천하는 방법이라 할 수 있다(Albadvi and Shahbazi, 2009; Liu *et al.*, 2011; Moon *et al.*, 2017; 박지희, 남기환, 2017). CBF와는 다르게 콘텐츠의 속성에 대한 정보 없이 유사 사용자들의 활동에 기반하여 콘텐츠를 추천하게 된다. 이러한 특성으로 인해 서비스 초기의 경우 사용자들의 활동정보가 부족하여 추천이 정상적으로 이루어지지 않는 Cold Start 문제가 발생할 수 있는 한계가 있다. CF의 Cold Start 문제를 해결하기 위하여 CBF와 CF를 함께 이용하는 Hybrid 방식이 널리 사용되고 있다. CF 사용자 활동정보 수집이 불가능할 경우 CBF를 이용하여 사용자 프로파일을 기반으로 추천을 하는 방식이다. LDA 기법을 이용하여 Hashtag를 추천하고자 하는 연구가 진행되었으며(Zhao *et al.*, 2016), Hashtag라고 명시되어있지는 않지만 이미지를 분석해서 연관된 특정 키워드를 추천해주거나 주요 관광지의 랜드마크를 파악하여 관련 정보를 제공하는 연구들이 진행

되었다(Cao *et al.*, 2015; Giovany *et al.*, 2017; Li *et al.*, 2016; Uricchio *et al.*, 2017; Zhang *et al.*, 2016). 이처럼 추천 시스템은 널리 사용되고 있으며 다양한 기법과 SNS와의 결합 등으로 그 사용의 범위가 점차 넓어지고 있다. 다만 Hashtag 추천에 있어서는 아직까지 진행된 연구가 많지 않다. 본 연구에서는 SNS 상의 Hashtag 추천을 위한 시스템을 설계하고 구현함으로써 사용자들에게 좀 더 쉽고 정확한 Hashtag를 제시하고자 한다.

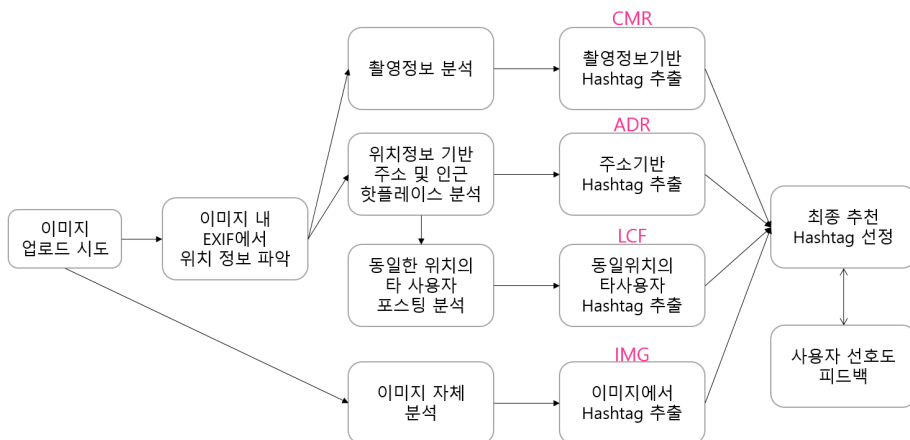
III. 시스템 설계 및 구현

3.1 시스템의 설계

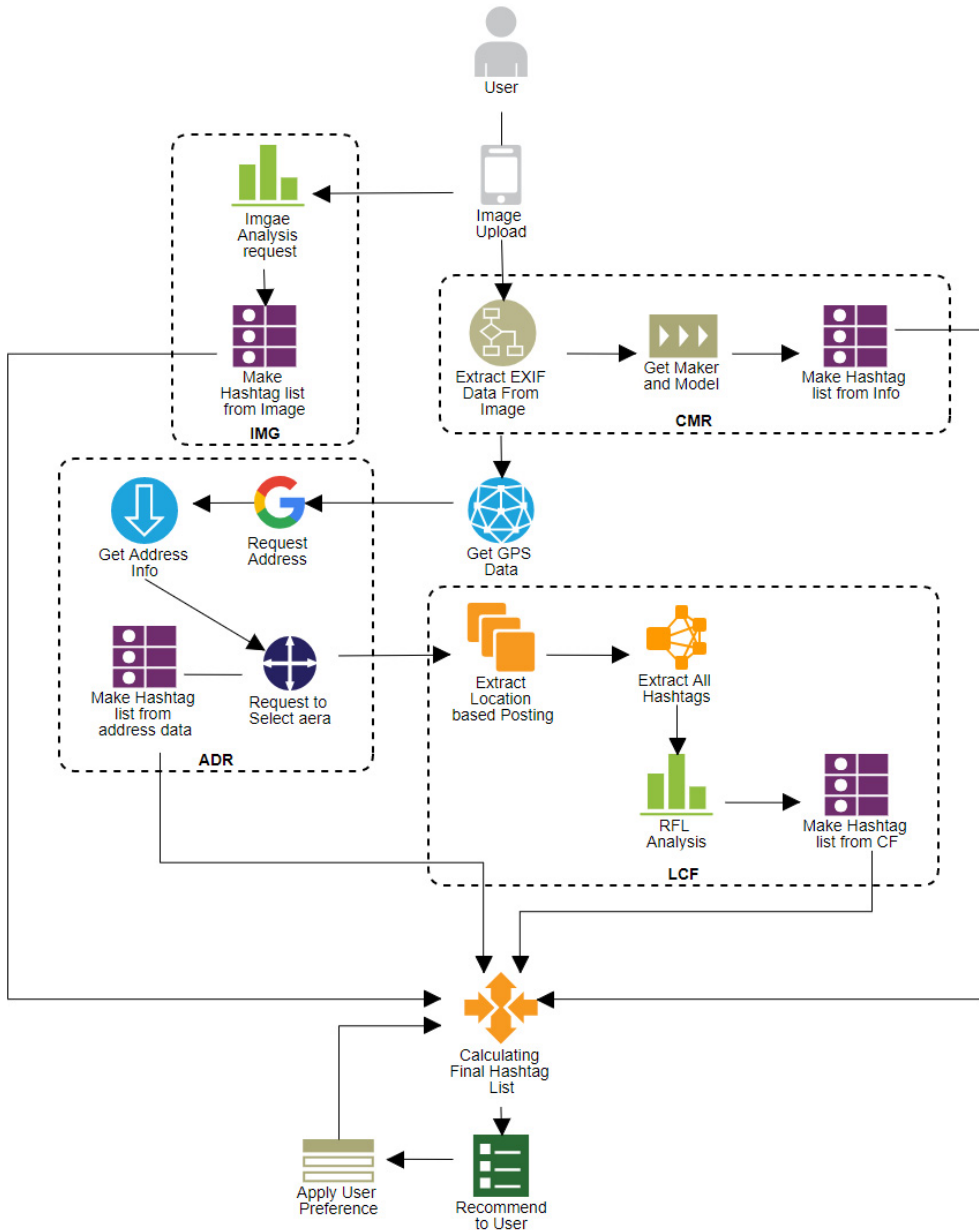
사용자들에게 최적의 Hashtag를 추천하기 위하여 본 연구에서는 <그림 1>과 같이 추천 프레임워크를 제안하였다. 본 모델에서는 크게 이미지의 정보를 파악하는 단계, 분석하는 단계, 추출하는 단계 그리고 최종 추천하는 단계로 나누어진다. 먼저 이미지 정보를 파악하는 단계에서는 사용자가 SNS 상에 이미지를 업로드하기 위해 파일을 첨부할 경우 첨부된 이미지의 EXIF 정보를 이용하여 위도와 경도에 대한 정보를 파악한다. 이미지가 촬영된 위도와 경도를 파악하게 되면 이 정

보를 이용하여 위치가 어디인지 인식 가능한 형태의 주소를 파악한다. 일반적으로 SNS는 사용자가 위치를 지정할 수 있게 되어 있으므로 사용자가 지정한 위치정보를 사용하거나 사용자가 지정하지 않을 경우 파악된 주소지 정보와 인근에서 가장 인기 있는 장소의 정보를 분석한다. 정확한 위치를 파악하게 된 경우 주소와 핫플레이스와 같은 주소 자체에 대한 Hashtag를 추출하게 된다. 또한 파악된 위치정보를 이용하여 타 사용자들이 동일한 위치에서 업로드 한 포스트를 검색할 수 있게 된다. 동일 위치에서 촬영한 이미지가 있는 포스트를 분석하면 포스트 내 포함된 타 사용자의 Hashtag를 수집하게 되고 수집된 정보를 기반으로 하여 추천 리스트를 작성할 수 있게 된다.

또한 이미지 자체에 대한 분석을 실시하여 이미지 내 포함된 오브젝트들을 추출하고 이를 Hashtag로 변환하여 추출한다. 마지막으로 위의 세 가지 차원에서 추출된 Hashtag를 정리하여 최종 추천 리스트를 작성한다. 또한 사용자가 어떠한 차원의 Hashtag를 선호하는지에 대하여 파악하고 개인화된 맞춤형 추천 시스템을 제공한다. <그림 2>는 제안 시스템에 대한 구성도를 나타낸다. 본 시스템은 촬영정보 기반, 주소 기반, 타사용자 기반, 이미지 기반을 이용한 추출단계를 포함하고 있다.



<그림 1> Hashtag 추천 프레임워크



〈그림 2〉 시스템 구성도

3.1.1 촬영정보기반 Hashtag 추출

사용자가 이미지를 등록할 경우 가장 먼저 시스템은 등록된 이미지 내의 EXIF 정보를 파악한다. EXIF 내 위치정보, 카메라의 제조사 정보, 모델명 등은 Hashtag로 사용이 가능하다. SNS 상에 카메라

모델명으로 검색할 경우 다양한 이미지를 볼 수 있는데 사용자들이 이를 Hashtag로 사용하고 있는 것을 알 수 있다. 또한 추출된 위치정보(위도, 경도)는 다음 단계인 주소기반 Hashtag 추출단계로 전달하게 된다.

3.1.2 주소기반 Hashtag 추출

앞선 단계에서 추출한 위도와 경도의 경우 위치에 대한 정보를 제공하지만 사람이 직관적으로 이해할 수 있는 정보는 아니다. 먼저 사람이 인식할 수 있는 주소 정보로 변환하여 사용해야 한다. Google Maps API를 이용하여 주소정보로 변환 시 Hashtag에 사용할 수 있는 공식적인 주소와 명칭, 건물명, 주와 도시 정보를 사용한다. 또한 사용자들이 업로드 하는 사진은 인기지역인 경우가 많으므로 인기지역 정보도 함께 받아와 Hashtag 추천에 사용한다.

변환 결과는 Hashtag로 사용할 수 있는 주소, 지명, 도시, 국가, 건물명 등의 정보를 획득할 수 있게 된다. 이러한 정보들은 Hashtag로서의 가치가 있는 정보임에 따라 향후 최종 추천에 사용될 수 있다. 다만 에펠탑과 같이 대형 구조물일 경우 촬영자의 위치와 실제 사용되는 Hashtag와의 차이가 있을 수 있다. 사진 촬영자의 위치가 항상 주요 지역과 일치하지 않을 수 있기 때문에 GPS 정보에 반경을 지정하여 이러한 문제를 해결하고자 하였다. 반경이 넓어질수록 이미지의 정확한 위치를 찾는데 문제가 생기는데 이를 개선하기 위하여 반경 내에서 일반적으로 가장 인기가 많은 지역을 파악하여 제공함으로써 정확도를 높일 수 있다. 인기있는 지역을 중심으로 추천을 시행하거나 몇 가지 대안을 제시한 뒤 사용자가 사진이 어떤 지역인지에 대하여 선택할 수 있는 옵션을 제공하는 방식을 채용할 수 있다. SNS의 경우 사용자가 직접 위치를 지정할 수 있는 기능이 있으므로 이를 이용하여 Hashtag 추천에 사용할 수도 있다. 위와 같은 절차를 거쳐 본 단계에서는 GPS 정보에서 추출한 주소 정보로 Hashtag 추천 리스트를 작성할 수 있다. 예를 들어 에펠탑을 배경으로 한 사진의 경우 Paris, France, Avenue Pierre Loti, Torre Eiffel 등이 추천 리스트에 올라갈 수 있다. 사용자가 특정지역을 선택한 경우 지정된 위치정보는 추천에 있어 가장 높은 우선순위를 배정받을 수 있으며 사용자가 특별히 지정하지 않은 경우 가장 인기 있는 지역

정보를 이용하여 추천 리스트에 있는 Hashtag들의 점수를 지정할 수 있다.

3.1.3 동일위치 타사용자 기반 Hashtag 추출

본 단계에서는 파악된 위치를 기반으로 타 사용자의 Hashtag를 추출하고 이를 이용하여 Tag 추천 리스트를 작성한다. 이 단계를 위해서는 먼저 추출된 위치를 기반으로 동일한 위치에서 작성된 포스팅을 검색해야 한다.

일반적인 SNS의 경우 특정 위치에서 업로드 된 포스팅을 일괄적으로 확인할 수 있다. 이 페이지 내 포함된 포스팅의 Hashtag를 추출하여 추천에 사용한다. 추출하는 정보는 Hashtag의 빈도와 Like 수 그리고 포스팅 게시일이다. Hashtag의 경우 추천을 위해 필수적인 정보이며 Like수의 경우 특정 Hashtag의 인기도를 파악하는데 도움을 줄 수 있다. 게시일 정보는 최근 이슈의 반영이나 특정 행사, 프로모션, 박람회, 축제 등의 정보를 반영하기 위하여 사용된다.

본 연구에서는 각각의 차원에 대한 점수를 산출하기 위하여 가치를 파악할 때 사용하는 RFM 기법을 응용하여 사용한다. RFM 기법은 특정 행위에 대한 빈도(Frequency), 최신성(Recency), 비용(Monetary)을 기반으로 한 가치 산출 기법이다. 본 연구에서는 이를 응용한 RFL 기법을 사용하였다. RFL은 Hashtag가 얼마나 최근에 업로드 되었는가(Recency), 얼마나 자주 업로드 되었는가(Frequency), 얼마나 많은 Like를 받았는가(Like)를 기반으로 산출하게 된다.

Frequency의 경우, 특정 위치에서 중복되게 출현하는 Hashtag가 존재할 수 있다. 이러한 Hashtag는 특정 위치를 가장 잘 설명하는 Tag라 할 수 있으며 Hashtag를 추천할 때 높은 점수를 부여하는 것은 타당하다고 할 수 있다. 본 연구에서는 이를 Frequency Score(FS)로 정의하여 분석하였다.

Frequency Score의 산출방법은 다음과 같다. 통상 RFM 기법에서 사용하는 5점 척도 대신 수식의 분모에 max 함수를 적용하여 각 점수의 최댓값이

1이 되도록 정규화 하였는데 이는 각 차원에 대한 대략적인 값을 직관적으로 이해하기 위한 것이다.

$$FS_{htag_i} = \frac{\text{count}(htag_i)}{\max\{\text{count}(htag_1), \text{count}(htag_2), \dots, \text{count}(htag_n)\}}$$

where n is number of distinct hashtags
 $i = 1, 2, \dots, n$
 $0 \leq FS \leq 1$

Frequency Score의 경우 가장 높은 빈도를 가진 Hashtag가 최고 점수인 1.0을 받게 된다. 그렇기 때문에 가장 높은 Hashtag의 빈도를 분모로 두고 각각 추출되는 Hashtag를 분자로 하여 FS를 계산한다.

<표 1> Frequency Score의 예시

Hashtag	Frequency	FS
Paris	400	1.0
Eiffeltower	385	0.96
France	373	0.93
Instatravel	328	0.82
Travel	311	0.78

<표 1>에서의와 같이 에펠탑에서 추출한 Hashtag 중 Paris의 빈도수가 400으로 가장 높다면 Paris의 FS는 1.0이 된다. 같은 방식으로 France의 빈도수는 373이며 이를 FS로 계산하면 0.93이 된다. Travel은 0.78이라는 점수를 가지게 된다.

Like수의 경우 Hashtag를 추천하는데 직접적으로 사용하는 것은 부적절하다. 이는 Like수의 특성상 연예인이나 여행전문작가 등과 같이 SNS 상에서 많은 사용자들로부터 관심을 받는 경우 높은 수의 Like수를 받을 수 있다. Like수의 경우 사용자 자체에 대한 인기척도도 반영되기 때문이다. 다만 Hashtag의 출현빈도가 비슷할 경우 Like수를 이용하여 우선순위를 정하는데 사용할 수 있으며 본 연구에서는 Like수를 일부 반영하였으며 이를 Like Score(LS)로 정의하였다.

Like Score의 산출방법은 아래와 같다.

$$LS_{htag_j} = \frac{\text{like_num}(htag_j)}{\max\{\text{like_num}(htag_1), \text{like_num}(htag_2), \dots, \text{like_num}(htag_n)\}}$$

where n is number of distinct hashtags
 like_num is sum of likes in specific hashtag
 $j = 1, 2, \dots, n$
 $0 \leq LS \leq 1$

최근에 자주 출현하는 Hashtag는 현재 특정 장소의 어떠한 특성을 나타내는 경우가 있다. 과거에 자주 언급된 Hashtag라도 최근에 언급이 거의 없다면 Hashtag가 의미하는 특성이 더 이상 그 장소에 존재하지 않을 수도 있으며 반대로 과거에 언급되지 않다가 최근 들어 반복적으로 언급되는 Hashtag는 새롭게 특정 위치에 등장한 특성을 설명해줄 수도 있다. 본 연구에서는 이를 Recency Score(RS)로 정의하였다. RS의 경우 <표 2>와 같이 특정 구간별로 점수를 부여할 수 있다.

Recency Score를 사용하는 이유는 최근의 트렌드를 반영하려는 것이기 때문에 시간이 길어질수록 그 의미가 낮아질 것으로 예상하여 이와 같은 방식을 사용하였으며 필요하다면 전체기간을 적절히 분할하여 각 단계별 점수를 부여하는 방식도 사용할 수 있다.

<표 2> Recency Score의 구간별 기준 및 점수

Criteria	RS	Condition
RS1	1.0	<= 1 day
RS2	0.9	<= 3 days
RS3	0.8	< 1 week
RS4	0.7	< 2 weeks
RS5	0.6	< 4 weeks
RS6	0.5	< 8 weeks
RS7	0.4	< 12 weeks
RS8	0.3	< 16 weeks
RS9	0.2	< 24 weeks
RS10	0.1	Over 24 weeks

각각의 차원에 대해 점수의 산출이 끝났다면 이를 이용하여 위치기반 Hashtag 추천 리스트를 작성해야 한다. 본 단계에서는 위치를 기반으로 3차원의 추천점수를 산출할 수 있으며 그 반영 비율은 필요에 따라 조정하여 사용할 수 있다. 각 차원에서 추출된 점수는 다음 수식을 통하여 위치기반 Hashatg 점수로 변환된 뒤 추천 리스트로 작성된다.

$$Final\ Score = w_1 \times Frequency\ Score + w_2 \times Recency\ Score + w_3 \times Like\ Score$$

위 수식은 앞서 분석한 FS, RS, LS에 대한 점수를 기반으로 각각의 가중치를 이용하여 최종 점수를 산출하는 것을 나타내고 있다.

본 연구에서는 각 차원에 대한 가중치 산출을 위하여 AHP(Analytic Hierarchy Process)를 이용한다(Saaty, 1994). RFM 기법을 사용할 경우 가중치를 선정하기 위한 절대적인 기준은 존재하지 않는다. 일부 연구자들은 동일한 비율을 사용하거나 사용자가 임의로 설정하는 형태 또는 상위 20%의 평균이나 정규분포를 이용하는 방법을 사용하기도 한다. 이 외에도 많은 연구자들이 적절한 가중치를 찾기 위하여 AHP를 이용하고 있으며 AHP는 전문가들을 대상으로 항목 간 상대비교를 통하여 중요도를 파악하고 이를 일관성지수를 통하여 검증함으로써 기존의 방법보다 더욱 정확한 값을 찾을 수 있다는 장점이 있다(곽승진, 배경재, 2011; 류권석, 김성수, 2017; 황현석 등, 2011). 본 연구에서는 Hashtag에 대한 이해도가 높은 연구자들과 인스타그램 팔로우 수가 300명이 넘는 사용자들을 대상으로 AHP 분석을 실시하였다. 분석결과 Frequency Score의 가중치는 0.501, Recency Score의 가중치는 0.302, Like Score의 가중치는 0.197로 나타났다. AHP 분석을 통하여 사용자들은 자주 사용되는 Hashtag를 가장 선호하며 그 다음으로 최근에 사용된 Hashtag를 선호하는 것을 알 수 있었다. 이에 본 연구에서는 AHP 분석 결과로 추출된 가중치를 이용하여 Final Score를 산출하였다. 본 연구에서는 AHP를 이용하여 초기 가중치를 설정하였으나 실제 시스템을 적용할 때에는 서비스 제공자에 따라 자유롭게 변경될 수 있으며 나아가 사용자의 추천 Hashtag 선택에 따라 선호하는 차원을 더욱 많이 추천할 수 있도록 개인화된 시스템을 제공할 수도 있을 것이다.

3.1.4 이미지 기반 Hashtag 추출

Hashtag는 이미지 내 포함된 사물이나 자연환경, 날씨 등의 정보도 사용할 수 있다. 이미지 분석을

위하여 Google Photos에 업로드 된 이미지를 기반으로 한 분석 도구인 Google Vision을 사용하도록 설계하였다. 분석을 위하여 사용자가 이미지를 업로드 할 때 EXIF 정보를 파악하는 것과 동시에 이미지를 Google Vision을 통하여 분석을 진행한다. 분석결과에는 이미지 내 사물이나 환경정보 등을 포함하고 있으며 유사도 또한 함께 제공 받을 수 있다. 분석 결과를 이용하여 이미지 기반의 추천 리스트를 작성한다.

3.1.5 최종 추천 리스트 작성

본 단계는 앞선 단계에서 추출한 4가지 차원 즉 촬영정보기반, 주소기반, 위치기반, 이미지 기반 Hashtag 리스트를 이용하여 사용자에게 제시할 최종 Hashtag 추천 리스트를 작성하는 단계이다. 사용자에게 4가지 차원을 기반으로 추천 리스트를 모두 보여주는 것이 가장 명확한 추천이 되겠지만 이 경우 너무 많은 수의 Hashtag를 추천하게 되어 불편함을 초래할 수 있다. 이에 각각의 정보에 대하여 적절한 반영비율이 필요하다. 각 차원별 중요도와 최종추천에 사용되는 수, 추출예시는 <표 3>과 같다.

<표 3> 각 차원별 분석결과 반영정보

차원	위치CF (LCF)	이미지 정보(IMG)	주소정보 (ADR)	촬영정보 (CMR)
중요도	높음	중간	낮음	낮음
추출 수	4~5개	2~3개	1~2개	1~2개
추출예시	파리여행 낭만 프랑스	하늘 비행기 잔디	파리 에펠타워 피에르르티	캐논 6D 아이폰 7 갤럭시 S8

일반적으로 Hashtag는 10개 이내로 사용하는 경우가 많다. 10개를 초과하게 되면 모바일 화면 상에서 3줄을 차지하게 된다. 스마트 디바이스의 특성상 전체 포스팅의 길이가 길지 않음에 따라 3줄이 넘어가게 되면 포스팅 내용의 반 이상을 차지 할 수도 있다. 이러한 이유로 사용자에게는 10개의 Hashtag를 추천 리스트로 제시함으로써 필요한 정보를 제공하도록 한다.

최종 추천 리스트의 작성은 차원별로 명확하게 명시된 숫자만큼 추출하는 방법이 있으며 다른 방법으로는 각 요인에 대해 가중치를 부여하여 사용자의 선호패턴에 따라 변화를 줄 수 있도록 하는 방법이 있다. 본 연구에서는 사용자의 선호패턴을 반영하기 위하여 각 차원별 가중치 개념을 도입하여 최종 추천 리스트를 작성할 수 있도록 하였다. 이 가중치는 각각의 차원에서 추천되는 Hashtag에 영향을 미치는 것이 아니라 각 차원별로 추천되는 Tag의 수에 영향을 줄 수 있다. 각 차원의 추천 가중치가 w_i 이고 추천 수가 n_i 일 경우 추천 결과 Hashtag의 수는 차원 별로 다음과 같이 산출될 수 있다.

$$(LCF, IMG, ADR, CMR) = (n \times w_1, n \times w_2, n \times w_3, n \times w_4)$$

초기 각 차원의 추천 가중치를 0.5, 0.3, 0.1, 0.1 이라 하고 추천 수가 10인 경우를 가정하면 추천 결과의 수는 각 차원 별로 (5, 3, 1, 1)이 되고 결과 리스트는 다음과 같은 집합으로 나타낼 수 있다.

{LCF1, LCF2, LCF3, LCF4, LCF5, IMG1, IMG2, IMG3, ADR1, CMR1}

이와 같이 각 차원별로 추천을 위한 개수를 설정하기 위해 가중치를 부여한 다음 최종적으로 추천 Hashtag 리스트를 작성하게 된다. <표 4>는 각 차원별 Hashtag의 우선순위를 어떻게 선정하는지에 대해 설명하고 있다.

<표 4> 각 차원별 우선순위 도출방법

차원	우선순위 도출방법
타사용자 기반	Hashtag의 출현횟수, LIKE수, 게시일 등을 복합적으로 고려
이미지분석 기반	분석결과 키워드의 이미지 유사성
주소정보 기반	지정된 반경 내 인지도역 순
촬영정보 기반	별도로 존재하지 않음

3.1.6 추천결과 제시 및 사용자 패턴 파악

이전 단계에서 작성된 Hashtag 최종 추천 리스트는 사용자가 포스팅 시 이미지를 업로드하면 분

석이 시작되고 사용자가 포스팅의 내용을 작성할 때 출력하게 된다. 사용자는 출력된 추천 리스트 중 자신이 원하는 Hashtag를 선택하게 되고 자동적으로 Hashtag의 형태로 포스팅 내 포함될 수 있다. 다만 사용자가 각 차원별로 선호하거나 기피하는 경우가 발생할 수 있다. 예를 들어 전문적인 카메라나 촬영 장비를 사용하는 사람들의 경우 촬영정보 기반의 Hashtag를 선호할 수 있으나 구형 스마트폰이나 저가의 장비를 이용한 사람 또는 사진 촬영에 크게 관심이 없는 사람들은 이 정보가 크게 의미가 없을 수 있다. 위와 같은 이유로 다양한 사용자들의 패턴을 반영하여 추천 리스트를 작성해야 할 필요가 있다. 사용자가 추천 리스트에서 특정 Hashtag를 선택하여 포스팅 내에 포함시킬 때 그 Hashtag의 차원을 선호한다고 생각할 수 있으며 지속적으로 선택하지 않은 차원의 Hashtag는 필요없는 정보라고 생각할 수 있다. 그러므로 사용자가 특정 Hashtag를 선택할 때 어떠한 차원에서 추출된 Hashtag인지 파악하고 차원별 반영비율을 수정하여 점차 사용자가 선호하는 정보를 더욱 많이 제공할 수 있도록 하였다. 사용자가 특정 Hashtag를 선택하면 선택된 차원의 가중치를 일정 수준만큼 증가시키고 다른 차원의 가중치를 감소시켜 점진적으로 사용자 맞춤형 추천 리스트가 되도록 하였다.

3.2 시스템의 구현


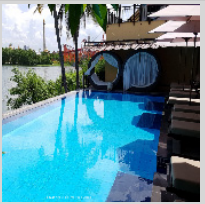
본 연구에서는 Python 언어를 사용하여 시스템을 개발하고 구현하였다. 또한 분석 과정 중 발생하는 정보들을 저장하기 위하여 MS-SQL을 사용하였다. 시스템을 개발한 환경은 Windows 10이며 Python IDE로 개발되었으며, MS-SQL의 경우 Windows Server 2013에서 서비스가 이루어졌다.

3.2.1 촬영정보 기반 Hashtag 추출 모듈

Hashtag 추천 시스템을 구현하기 위하여 EXIF의 정보를 파악하는 모듈을 구성하였다.

본 모듈은 Python의 Image Library를 사용하였으며 분석에 필요한 위치정보를 추출하였고 촬영 정보를 기반으로 한 Hashtag를 추출하기 위하여 촬영된 카메라의 제조사와 모델을 추출하여 추천 리스트를 작성하였다. <표 5>는 EXIF 정보파악 모듈을 이용하여 각 카메라의 제조사와 세부모델명 그리고 다음 단계에서 이용할 수 있는 GPS 정보를 포함하고 있는 것을 확인할 수 있다.

<표 5> EXIF 추출결과 및 카메라 정보 추천 Hashtag

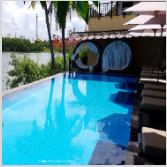
Image		
	Maker	PENTAX
Model	Ist DS	SM-G930S
GPS_Data	49.195, -123.180	16.045, 108.227
Hashtag	PENTAX Ist DS	Samsung

3.2.2 주소기반 Hashtag 추출 모듈

본 모듈은 GPS 정보를 사용자가 인식할 수 있는 주소로 변경하고 이를 기반으로 Hashtag를 추출하는 모듈이다. Google Map API에 변환을 요청하면 사람이 인식할 수 있는 주소로 변경된다. GPS의 위치를 중심으로 반경 100m를 설정한 뒤 반경 내 포함된 위치정보를 추출함으로써 사용자가 자신의 위치를 선택할 수 있도록 하는 방식과 반경 내 가장 인기있는 지역을 선택하여 추천하는 방식을 사용할 수 있다.

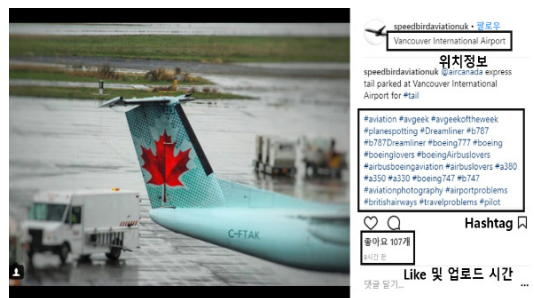
<표 6>은 다낭의 한 스파에서 촬영한 이미지의 GPS 정보를 주소기반으로 변경한 것과 주소를 기반으로 주변 반경 내 핫스팟을 보여주고 있다. 실제로 이 스파가 주변 지역 내에서 가장 인기있는 지역이라는 것을 확인할 수 있으며 이 경우 추천되는 Hashtag로 사용된다.

<표 6> 주소기반 Hashtag 추천 결과

Image	
Location	528 2 Tháng 9, Hòa Cường Bắc, Hải Châu, Đà Nẵng, Vietnam
Extracted Address & Hot Spot	Da Nang ZEN SPA- ZEN CORPORATION The Blossom Resort Danang Etron Smart Home(LEVITON USA) Water lily villa
Hashtag (Top Hot Spot)	ZEN SPA- ZEN CORPORATION

3.2.3 동일위치 타 사용자 기반 Hashtag 추출 모듈

본 단계에서는 이전 단계에서 파악된 사용자가 업로드 한 이미지의 위치를 기반으로 하여 유사한 사용자들의 Hashtag를 파악하고 분석하여 Hashtag를 추천하는 모듈이다. 구현한 시스템은 특정 위치에 업로드된 모든 포스트를 파악하여 Hashtag를 추출한다. Hashtag의 경우 #으로 분류되어 있기 때문에 쉽게 추출이 가능하다.



<그림 3> SNS에서 추출가능한 정보

<그림 3>은 SNS 상에서 획득할 수 있는 정보를 나타내고 있다. 포스트 상단에는 현재 포스트의 위치를 확인할 수 있다. 위 그림의 경우 밴쿠버 국제공항에서 업로드 된 포스트임을 확인할 수 있다

본문에서는 사용자가 작성한 Hashtag 리스트를 확인할 수 있다. 아래쪽에서 Like의 수와 업로드된 날짜를 확인할 수 있다. 이러한 정보를 이용하여 추천 시스템은 Hashtag와 함께 Like수, 게시일 정보를 함께 추출하여 복합적인 분석을 실시할 수 있도록 한다. 본 단계에서는 앞서 살펴본 RFL 기법을 이용하여 각각의 Hashtag에 대한 분석과 추천지수를 산출하게 된다. 수집된 Hashtag의 FrequencyScore를 분석하기 위하여 SQL명령문을 이용하여 Hashtag를 정렬하였다. 각 Hashtag의 출현빈도를 분석하기 위하여 count 기능과 Aliases 기능을 이용하였다.

유사한 형식으로 각 Hashtag에 대한 Like수를 파악할 수 있다. Like수는 앞서 빈도수를 파악하는 것과는 다르게 각각의 Hashtag별 Like수의 합계를 구하는 것으로 count 명령어 대신 sum 명령어로 총 합계를 구하였다.

RecencyScore의 경우 각 Hashtag가 포함된 포스팅의 업로드 날짜를 이용하였다. 각각의 포스팅은 업로드 날짜가 명시되어 있고 동일한 포스팅에서 사용한 모든 Hashtag는 같은 RecencyScore를 갖게 된다.

만약 다양한 포스팅에서 여러 번 나타난 Hashtag의 경우 가장 최근의 Hashtag에 대한 RecencyScore를 이용하는 방법과 여러 번 나타날 때마다 각각 부여된 점수를 모두 합산 반영할 수 있다. 본 연구에서는 최근 들어 빈도가 높아진 Hashtag를 파악하기 위하여 반복하여 나타나는 Hashtag들의 RS값을 모두 더하여 최종 RecencyScore를 산출하였다. <그림 4>는 CF점수를 산출하는 과정을 일부 제시하고 있다.

```
CF기반 추천 =====
tag = paris num = 25
sum like = 972.0
tag = eiffeltower num = 14
sum like = 587.0
tag = toureiffel num = 13
sum like = 414.0
tag = france num = 12
sum like = 419.0
tag = travel num = 12
sum like = 633.0
tag = picoftheday num = 8
sum like = 234.0
tag = love num = 7
sum like = 486.0

해시태그 : Paris
CF기반점수 : 1000.0
해시태그 : eiffeltower
CF기반점수 : 583.248089359
해시태그 : toureiffel
CF기반점수 : 468.306878307
해시태그 : france
CF기반점수 : 491.678424456
해시태그 : travel
CF기반점수 : 505.12345679
해시태그 : picoftheday
CF기반점수 : 315.502649503
해시태그 : Love
CF기반점수 : 281.428571429
해시태그 : instatravel....
```

<그림 4> CF점수 연산과정

3.2.4 이미지 기반 Hashtag 추출 모듈

본 단계에서는 이미지 자체를 분석한다. 분석을 위해 Google Vision을 이용하였고 사용자가 업로드한 이미지를 전송하면 서버에서 분석하고 그 결과를 Python의 List Type으로 반환한다. 반환되는 정보는 추출된 이미지 분석 결과와 유사도이다. 추천을 위하여 반환되는 정보를 이용하여 가장 유사한 분석 결과 순으로 정렬하고 최종 추천에 사용하도록 한다. <그림 5>은 이미지 기반 추천점수를 산출하는 과정 중 일부를 나타내고 있다.

```
이미지 기반 추천 =====
sky
0.932791411877
landmark
0.910039365292
tree
0.766279041767
tower
0.748991429806
tourism
0.701344430447
monument
0.685948073864
grass
0.538957777023
outdoor_structure
0.58445235014
vacation
0.545961260796
recreation
0.544924139977
----- 이미지 기반 결과 DB 업데이트 ----
DB Update..... Ok
===== 이미지 기반 추천리스트 작성 종료=====
```

<그림 5> 이미지 기반 추천점수 산출 과정

3.2.5 최종 추천 및 사용자 패턴 파악 모듈

<표 7> 각 이미지별 최종 추천 Hashtag List

			
LCF1	Paris	Vancouver	Vietnam
LCF2	Eiffeltower	Yvr	Hoian
LCF3	Travel	Travel	Travel
LCF4	Toureiffel	Canada	Instatravel
LCF5	France	Aviation	Wanderlist
IMG1	Sky	Airline	Night
IMG2	Landmark	Airport	Light
IMG3	Tree	AirportApron	Lighting
ADR1	Eiffel Tower	Vancouver Aport	Hoi An
CMR1	Canon	Pentax Corp	Samsung

본 모듈에서는 앞서 수집한 4가지 단계의 Hashtag에 대하여 사용자에게 추천하는 모듈이다. 이 단계에서는 각 단계에서 추출된 Hashtag를 어느 정도 사용자에게 제시할 것인지에 대한 연산을 수행한다. 또한 추천된 결과에 대해 사용자가 어떠한 단계의 결과를 선호하는지 파악하여 다음 추천에서 이를 반영한 추천이 이루어지도록 하였다. <표 7>은 이미지별 최종 추천 Hashtag를 나타낸다.

IV. 시스템 성능 테스트

4.1 테스트 방법

구현한 시스템의 성능을 검증하기 위하여 세 장의 이미지를 토대로 테스트를 진행하였다. 이미지 기반 SNS로 오랜 기간 사용자들에게 사랑받고 있는 Flickr의 Hashtag 추천 기능과 전문사진 SNS로 인기를 끌고 있는 500px 서비스의 Hashtag 추천 기능을 이용하여 제안하는 시스템과 비교분석 하였다. 또한 추천된 결과를 이용하여 SNS의 사용자들에게 조사를 실시하여 추천된 Hashtag가 실제 사용에서 유용한지에 대한 사용자 평가를 진행하였다.

4.2 타 서비스와 Hashtag 추출 비교

테스트를 위하여 이미지 세 장을 비교대상이 되는 Flickr, 500px 서비스로 업로드 하여 추천되는 Hashtag에 대한 정보를 분석하였다. <표 8>에서 확인할 수 있듯이 500px의 경우 추천된 개수가 12개로 가장 많지만 추천결과 중 다수가 관계없는 내용을 추천하고 있는 것을 알 수 있다. 반면 Flickr의 경우 추천되는 수는 부족하나 이미지에 대한 일부 설명을 포함한 Hashtag를 추천하고 있다. 제안 시스템의 경우 이미지에 대한 다양한 정보를 제공하고 있다.


<표 9>에서는 밴쿠버공항 이미지에 대한 타 서비스와의 Hashtag 추천 결과를 확인할 수 있다. 본 이미지에서 500px의 경우 앞선 에펠탑 이미지와

동일하게 총 12개의 Hashtag를 추천하였다. 하지만 추천 수에 비해 대부분 의미 없거나 관련성이 적은 Hashtag를 추천하고 있음을 알 수 있다. Flickr의 경우 위치정보에 기반한 Hashtag 추천 결과를 보여주는 것으로 보이나 추천되는 수가 3개로 상대적으로 적음을 알 수 있다. 반면 본 연구

<표 8> 에펠탑 이미지에 대한 추천 결과 비교

	Proposed System	500px	Flicker
1	Paris	Baton	Grass
2	Eiffeltower	Red Dress	Tower
3	Travel	Pagoda	
4	Toureiffel	White Dress	
5	France	Archer	
6	Sky	Dress	
7	Landmark	Victory	
8	Tree	Dance	
9	Eiffel Tower	Monk	
10	Canon	Temple	
11		Park	
12		Dancer	

<표 9> 밴쿠버공항 이미지에 대한 추천 결과 비교

	Proposed System	500px	Flicker
1	Vancouver	Interstate	Richmond
2	Yvr	Car	British Columbia
3	Travel	Transportation	Canada
4	Canada	Expressway	
5	Aviation	Drive	
6	Airline	Freeway	
7	Airport	Driving	
8	Airport Apron	Semaphore	
9	Vancouver Airport	Taxi	
10	Pentax Corp.	South Beach	
11		Highway	
12		Skyline	

에서 제안하는 추천 시스템의 경우 캐나다의 밴쿠버 공항에 대한 Hashtag 뿐 아니라 공항과 관련된 일반적인 Hashtag를 적절하게 추천해주고 있음을 확인할 수 있다.

<표 10> 호이안 이미지에 대한 추천 결과 비교

	Proposed System	500px	Flicker
1	Vietnam	Night	People
2	Hoian	Festival of Light	
3	Travel	NightLife	
4	Instatravel	New year eve	
5	Wanderlist	Lighted	
6	Night	Nightshot	
7	Light	Firework	
8	Lighting	Light Show	
9	Hoi An	City Light	
10	Samsung	Nightout	
11		Night Shot	
12		City at Night	

<표 10>은 호이안 이미지에 대한 분석 결과를 보여주고 있다. 500px의 경우 12개의 Hashtag를 추천해주었고 야간사진이라는 것에 초점을 맞춰 적절한 Hashtag를 앞선 이미지들보다 더 정확하게 추천하고 있다. 반면 Flicker의 경우 People이라는 하나의 Hashtag만을 추천하여 만족스럽지 못한 결과를 보여주고 있다. 제안 시스템의 경우 Instravel과 같은 인스타그램에서 사용하는 전용 Hashtag와 함께 호이안에 대한 적절한 Hashtag를 추천하는 것을 확인할 수 있다.

추가적으로 실시된 실험에서 <표 11>은 실내에서 촬영한 음식사진에 대한 추천 시스템별 결과를 보여주고 있다. 제안한 시스템의 경우 사진에 대한 정보를 잘 나타내고 있으나 500px의 경우 불필요한 Hashtag를 추천해주고 있는 것을 볼 수 있다. 반면 Flicker의 경우 비교적 정확한 Hashtag를 추천하고 있으나 추천 개수가 2개 밖에 되지 않아 제안

시스템 대비 충분한 정보를 제공하지 못하는 것으로 나타났다.

<표 11> 음식 이미지에 대한 추천 결과 비교

	Proposed System	500px	Flicker
1	딩케슈니첼	Liqueur	Food
2	독일마을	Oktoberfest	Meal
3	남해	Night market	
4	슈니첼	Mexican	
5	남해여행	Caribbean	
6	Food	Oktoberfest	
7	Brunch	Cocktail	
8	Breakfast	Nightlife	
9	딩케슈니첼	Festival of Lights	
10	Samsung	Southeast	
11		Casino	
12		Garnish	

4.3 추천 결과 평가

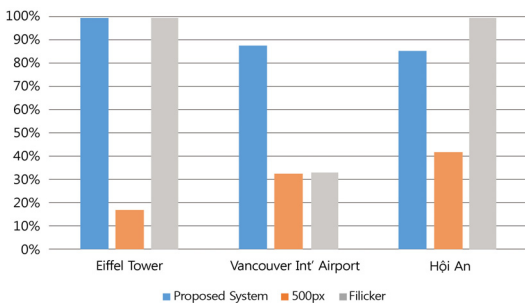
제안 시스템의 추천 결과가 수용할 만한 수준인지 확인하기 위하여 사용자 평가를 진행하였다. 평가를 위해 현재 SNS를 사용하고 있는 사용자가 어떠한 Hashtag를 선택하는지에 대한 분석을 진행하였다. 분석을 위해 사용된 이미지는 에펠탑, 밴쿠버공항, 호이안으로 다양한 국가의 응답자들이 제대로 인지하고 평가할 수 있도록 널리 알려진 장소를 사용하였다. 사용자 분석은 온라인과 오프라인에서 동시에 진행되었으며 대한민국을 포함하여 독일, 멕시코, 미국, 세르비아, 이란, 이스라엘, 카메룬, 코트디부아르, 키르기스스탄, 타지키스탄, 폴란드, 잠비아, 필리핀, 불가리아 등 총 15개국의 사람들이 참여하였다.

<표 12>에서 확인할 수 있듯이 전체 응답자 중 남성이 64%, 여성이 35%를 차지하고 있으며 SNS의 주사용자인 20대에서 40대가 대부분으로 본 연구의 평가 집단으로서 적절하다고 판단된다.

〈표 12〉 응답자의 인구통계학적 정보

기준	세부항목	인원(명)	비율(%)
성별	남성	136	64.2
	여성	75	35.3
	미응답	1	0.5
나이	20대	64	30.2
	30대	65	30.7
	40대	73	34.4
	50대 이상	8	3.8
	미응답	2	0.9

추천 시스템의 평가에는 정확률(Precision)과 재현율(Recall)이 일반적으로 사용된다. Precision이란 결과의 정확도를 확인하는 척도로 결과로 제시된 정보에서 얼마나 정확한 정보가 포함되어 있는지에 대한 기준이다. Recall의 경우 결과로 제시된 정보가 대상이 되는 적합한 정보의 전체 수와 비교해서 얼마나 정확한지에 대하여 나타내는 비율을 의미한다(Salton and McGill, 1983). Precision과 Recall은 많은 연구에서 시스템의 성능 평가에 사용되었다(Barragáns-Martínez et al., 2010; Cho et al., 2007; Fard et al., 2013; Jang et al., 2008; Lathia et al., 2009; Paranjape-Voditel and Deshpande, 2013; Shih and Liu, 2008; 박세화, 박석, 2016).



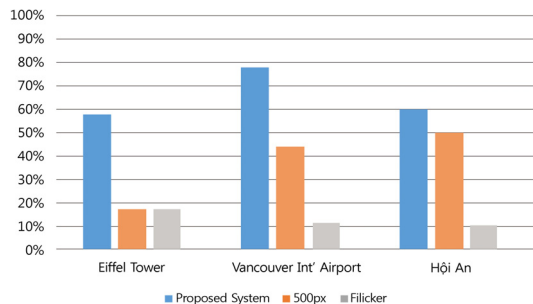
〈그림 6〉 Precision 분석 결과

〈그림 6〉은 앞선 단계에서 추출한 Hashtag를 기반으로 사용자 평가 후 Precision에 대한 결과이다. 에펠탑에 대한 분석 결과 제안 시스템과 Flickr의 결과가 높은 값을 보였다. 제안 시스템의 경우

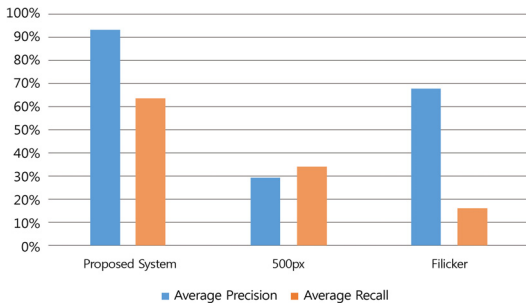
여러 개의 Hashtag를 추천하여 모두 적절한 Hashtag로 확인된 반면 Flickr는 2개의 Hashtag를 추천하고 이 2개가 적절한 것으로 판단되어 Precision이 100%로 나오게 되었다. 밴쿠버공항의 경우 제안 시스템의 Precision값이 가장 높게 나타났고 500px과 Flickr는 제안 시스템보다 낮은 33%를 나타냈다. 호이안의 경우 제안 시스템은 86%의 값을 보이나 Flickr는 100%를 나타내고 있는데 이는 Flickr가 한 개의 Hashtag만을 추천했으며 추천한 1개의 결과가 적절한 것으로 판단되어 100%로 나타난 것이다.

Precision의 경우 한 개의 추천 결과에서 하나의 정답만 맞추게 되어도 100%를 획득할 수 있어 Flickr는 적은 수의 추천을 통해 높은 Precision 비율을 획득하였다. 그러나 Recall 비율은 매우 낮게 나와 시스템 성능에 문제가 있는 것을 확인할 수 있다.

Recall에 대한 결과는 〈그림 7〉과 같다. 세 경우 모두 테스트 결과에서 제안 시스템이 가장 높은 성능을 보여주는 것을 확인할 수 있다. Recall은 사용자가 선택한 Hashtag 중 얼마나 많은 Hashtag를 시스템이 제시했는지에 대한 것으로, 제안 시스템은 적절한 추천 수와 함께 채택률을 동시에 만족했으므로 결과적으로 높은 점수를 받을 수 있었다. 이전 단계에서 높은 Precision값을 획득한 Flickr의 경우 Recall에서는 매우 낮은 값을 보여주었는데 이는 적은 수의 추천 결과를 제시하였기 때문에 정확하게 추천된 정답의 수 또한 매우 부족하기 때문이다.



〈그림 7〉 Recall 분석 결과



<그림 8> Precision과 Recall의 평균값

<그림 8>은 Precision과 Recall의 평균값을 나타내고 있는데 결과에서 확인할 수 있듯이 제안 시스템이 현재 서비스 되고 있는 다른 시스템에 비해 매우 높은 결과를 보여준다. 분석 결과를 종합해 보면 제안 시스템이 대부분의 테스트 환경에서 높은 평가를 받았으며 기존 시스템 대비 성능이 향상되었음을 확인할 수 있다.

V. 결론 및 향후연구

본 논문은 SNS 상에서 사용자가 업로드 하는 이미지를 기반으로 적절한 Hashtag를 추천하는 연구에 관한 것이다. 본 연구를 통하여 사용자들이 자신의 이미지를 업로드 할 때 보다 쉽고 간단한 방법으로 Hashtag를 추가할 수 있도록 하고자 하였다.

본 연구에서는 Hashtag를 추천하기 위하여 크게 4가지 기반의 분석을 실시하였다. 각각의 분석은 촬영정보에 기반한 분석(CMR), 주소에 기반한 분석(ADR), 동일 위치의 타사용자에 기반한 분석(LCF), 이미지 자체에 기반한 분석(IMG)으로 설계하였다. 또한 본 연구에서는 위의 형태로 구성된 시스템을 설계하고 이를 구현하였다. 구현을 위하여 Python을 이용하였으며 정보를 저장하기 위한 MS-SQL Server를 함께 사용하였다. EXIF 정보를 파악하기 위하여 Python Image Library를 이용하였으며 이를 통해 촬영정보와 함께 GPS 정보도 확보할 수 있었다. 확보된 촬영정보는 촬영정

보 기반 Hashtag 리스트로 분류되고 GPS 정보는 Google Map의 데이터를 이용하는 Google Cloud를 이용하여 사용자가 인식할 수 있는 주소체계로 변경하였으며 이를 통해 주소정보에 기반한 Hashtag 추천 리스트를 확보할 수 있었다.

위치정보를 이용하면 SNS 상에 업로드 된 사용자의 정보를 파악할 수 있다. 이러한 정보를 분석하기 위하여 SNS 상에서 특정 위치에 업로드 된 Hashtag 정보를 수집하였고 Hashtag의 빈도, 최신성, Like수 등을 고려하여 사용자에게 가장 적합한 Hashtag를 추천하고자 하였다. 빈도와 최신성의 경우 특정 Hashtag가 자주 사용되고 반복된다는 것을 의미하며 Like수의 경우 인기 게시물에 더욱 많은 Like가 있으므로 이를 반영하는 것이 의미가 있을 것이다. 이들 정보를 RFM 기법을 응용한 RFL이라는 개념으로 분류하고 이를 기반으로 하여 각 Hashtag에 대한 추천 지수를 파악하고 사용자 기반의 Hashtag 추천 리스트를 작성하였다. 또한 이미지를 분석하기 위해 Google Vision을 이용하였다. 사용자가 업로드 한 이미지를 Google Vision Server로 전송하고 분석된 결과를 이용하여 이미지 기반의 Hashtag 추천 리스트를 작성하였다. 4가지의 차원에서 추출한 Hashtag 리스트를 이용하여 사용자에게 적합한 최종 Hashtag 추천 리스트를 제시하였고 사용자의 선택에 따라 추천 비율을 변경할 수 있도록 하는 개인화 방안을 설계 및 구현하였다.

본 연구에서 구현한 시스템이 기존 시스템에 비해 개선되었는지 확인하기 위하여 Hashtag 추천 기능을 포함하고 있는 타 시스템과의 성능 테스트를 실시하였다. 오랫동안 이미지 기반 SNS의 최대 서비스였던 Flickr와 모바일 전용 이미지 SNS인 500px을 이용하여 제안 시스템과 함께 동일한 이미지에 대한 분석을 진행하였다. 제안 시스템의 검증을 위해 15개국 210명의 SNS 사용자를 대상으로 Precision과 Recall 점수를 이용한 성능 평가를 진행하였고 평가 결과 제안 시스템이 기존 시스템 대비 개선된 성능을 보여주었다.

우수한 사용자 평가의 결과에도 불구하고 본 연구는 다음과 같은 한계를 가진다. 먼저 타 사용자의 Hashtag 추출에 있어 일부 사용자는 이모티콘을 적용한 Hashtag를 사용하였는데 이러한 정보는 본 연구에서 고려하지 않았다. 또한 Hashtag는 일반적인 정보뿐 아니라 특정 상황에서 독특한 사용자 감성을 표현하는 수단으로 사용되기도 하는데 이러한 정보는 RFL 추천에서 높은 점수를 획득하기 어려워 추천이 되지 않는다는 문제가 있다. 또한 반영비율을 선정하는데 있어 관련 전문가 그룹과 유저그룹을 함께 사용하였다는 한계가 있다. 그리고 최근 들어 SNS에서 복수 개의 이미지를 하나의 포스팅 내에 첨부할 수 있도록 하는 움직임이 있는데 이 경우 추천의 정확도가 떨어질 수 있다. 마지막으로 추천결과 평가에 있어 유명 관광지 세 장소만을 이용했다는 한계가 존재한다. 인물, 실내, 실외, 음식, 동물 등 다양한 포스팅에 대한 평가가 향후에 추가적으로 진행되어야 할 것이다.

참 고 문 헌

- [1] 광승진, 배경재, “전자책 단말기의 사용성 평가에 관한 실험적 연구”, *정보관리학회지*, 제28권, 제3호, 2011, pp. 313-333.
- [2] 김세진, 이상돈, “SNS 환경에서 동시출현 해시태그 분석 기법”, *한국콘텐츠학회 종합학술대회논문집*, 2014, pp. 223-224.
- [3] 남민지, 이은지, 신주현, “인스타그램 해시태그를 이용한 사용자 감정 분류 방법”, *멀티미디어학회논문지*, 제18권, 제11호, 2015, pp. 1391-1399.
- [4] 류권석, 김성수, “모바일 쇼핑 구매요인의 우선순위 분석: 모바일 라이프 스타일과 소비자 의사결정 성향을 중심으로”, *서비스경영학회지*, 제18권, 제2호, 2017, pp. 197-223.
- [5] 류동성, 정우근, 조환규, “격자 기반의 디지털 사진 시각화와 계층적인 클러스터링 방법”, *정보과학회논문지: 컴퓨팅의 실제 및 레터*, 제16권, 제5호, 2010, pp. 616-620.
- [6] 문지훈, 유제혁, 김용성, 황인준, “SNS 해시태그 분석을 통한 목적별 여행지 분류 기법”, *한국정보과학회 학술발표논문집*, 2015, pp. 121-123.
- [7] 박세화, 박석, “체류시간을 고려한 여행 일정 추천 시스템”, *정보과학회 컴퓨팅의 실제 논문지*, 제22권, 제1호, 2016, pp. 38-43.
- [8] 박지희, 남기환, “추천 수량과 재 추천을 고려한 사용자 기반 협업 필터링 추천 시스템”, *Information Systems Review*, 제19호, 제2호, 2017, pp. 71-94.
- [9] 안상욱, *해시태그 용어로 보는 IT*, BLOTTER.NET, 2015.
- [10] 장문종, 이웅희, 김영훈, “해킹된 트위터 사용자의 해시태그 분석 및 검출 방법”, *한국정보과학회 학술발표논문집*, 제6호, 2017, pp. 248-250.
- [11] 장철진, 주영호, 조환규, “EXIF정보를 이용한 디지털 사진 클러스터링 알고리즘”, *한국정보과학회 가을 학술발표논문집*, 제33권, 제2호, 2006, pp. 442-447.
- [12] 전민아, 전상훈, 노승민, 황인준, “트위터 해시태그와 트윗 분석을 통한 음악 추천 기법”, *한국정보과학회 학술발표논문집*, 2013, pp. 1257-1259.
- [13] 제성관, 이근동, 오원근, “부가정보를 이용한 자동 사진분류방법”, *한국HCI학회 학술대회*, 2011, pp. 573-575.
- [14] 최재용, 원지섭, 이수안, 김진호, “스마트폰에서 계층 모델 기반의 사진 자동 분류 및 사진 탐색기”, *정보과학회논문지: 컴퓨팅의 실제 및 레터*, 제18권, 제9호, 2012, pp. 647-653.
- [15] 최홍선, 이강희, 임광혁, 김수균, “EXIF 메타데이터를 활용한 웹 기반 사진 자동분류”, *한국컴퓨터정보학회 학술발표논문집*, 제20권, 제2호, 2012, pp. 301-302.
- [16] 황현석, 이상훈, 김수연, “Fuzzy AHP를 활용한

- 스마트폰 선택 및 이용 평가요인에 관한 연구”, *한국산업정보학회논문지*, 제16권, 제4호, 2011, pp. 107-117.
- [17] Alarifi, A., M. Alsaleh, and A. Al-Salman, “Twitter turing test: Identifying social machines”, *Information Sciences*, Vol.372, 2016, pp. 332-346.
- [18] Albadvi, A. and M. Shahbazi, “A hybrid recommendation technique based on product category attributes”, *Expert Systems with Applications*, Vol.36, No.9, 2009, pp. 11480-11488.
- [19] Arnaboldi, V., M. Conti, M. La Gala, A. Passarella, and F. Pezzoni, “Ego network structure in online social networks and its impact on information diffusion”, *Computer Communications*, Vol.76, 2016, pp. 26-41.
- [20] Barragáns-Martínez, A. B., E. Costa-Montenegro, J. C. Burguillo, M. Rey-López, F. A. Mikic-Fonte, and A. Peleteiro, “A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition”, *Information Sciences*, Vol.180, No.22, 2010, pp. 4290-4311.
- [21] Bennett, S., *The History of Hashtags in Social Media Marketing*, 2014, Available at <http://www.adweek.com/digital/history-hashtag-social-marketing/>.
- [22] Cao, J., Y. Zhao, X. Lai, M. E. H. Ong, C. Yin, Z. X. Koh, and N. Liu, “Landmark recognition with sparse representation classification and extreme learning machine”, *Journal of the Franklin Institute*, Vol.352, No.10, 2015, pp. 4528-4545.
- [23] Cheng, Z. and J. Shen, “On very large scale test collection for landmark image search benchmarking”, *Signal Processing*, Vol.124, 2016, pp. 13-26.
- [24] Cho, J. C. J., K. K. K. Kwon, and Y. P. Y. Park, “Collaborative filtering using dual information sources”, *IEEE Intelligent Systems*, Vol.22, No.3, 2007, pp. 30-38.
- [25] Fard, K. B., M. Nilashi, M. Rahmani, and O. Ibrahim, “Recommender system based on semantic similarity”, *International Journal of Electrical and Computer Engineering*, Vol.3, No.6, 2013, pp. 751-761.
- [26] Giovany, S., A. Putra, A. S. Hariawan, and L. A. Wulandhari, “Machine learning and SIFT approach for indonesian food image recognition”, *Procedia Computer Science*, Vol.116, 2017, pp. 612-620.
- [27] Huang, H. and W. Fang, “Metadata-based image watermarking for copyright protection”, *Simulation Modelling Practice and Theory*, Vol.18, No.4, 2010, pp. 436-445.
- [28] Jang, M. K., H. K. Kim, and J. K. Kim, “An extended content-based procedure to solve a new item problem”, *Journal of Intelligence and Information Systems*, Vol.14, No.4, 2008, pp. 201-216.
- [29] Lathia, N., X. Amatriain, and J. M. Pujol, “Collaborative filtering with adaptive information sources”, *IJCAI Workshop on Intelligent Techniques for Web Personalisation and Recommender Systems*, Pasadena, 2009.
- [30] Li, K., F. Wang, and L. Zhang, “A new algorithm for image recognition and classification based on improved Bag of features algorithm”, *Optik-International Journal for Light and Electron Optics*, Vol.127, No.11, 2016, pp. 4736-4740.
- [31] Liu, D. R., C. H. Lai, and H. Chiu, “Sequence-based trust in collaborative filtering for document recommendation”, *International Journal of Human-Computer Studies*, Vol.69, No 9, 2011, pp. 587-601.
- [32] Liu, X., L. Zhang, M. Li, H. Zhang, and D. Wang, “Boosting image classification with LDA-based feature combination for digital photograph management”, *Pattern Recognition*, Vol.38, No.6,

- 2005, pp. 887-901.
- [33] Moon, H. S., J. H. Yoon, I. Y. Choi, and J. K. Kim, "An exploratory study of collaborative filtering techniques to analyze the effect of information amount", *Asia Pacific Journal of Information Systems*, Vol.27, No.2, 2017, pp. 126-138.
- [34] Moorley, C. R. and T. Chinn, "Nursing and twitter: Creating an online community using hashtags", *Collegian*, Vol.21, No.2, 2014, pp. 103-109.
- [35] Paranjape-Voditel, P. and U. Deshpande, "A stock market portfolio recommender system based on association rule mining", *Applied Soft Computing*, Vol.13, No.2, 2013, pp. 1055-1063.
- [36] Pazzani, M. and D. Billsus, "Learning and revising user profiles: The identification of interesting web sites", *Machine Learning*, Vol.27, No.3, 1997, pp. 313-331.
- [37] Saaty, T. L., "How to make a decision: The analytic hierarchy process", *Interfaces*, Vol.24, No.6, 1994, pp. 19-43.
- [38] Salton, G. and M. J. McGill, *Introduction to Modern Information Retrieval*, New York: McGraw-Hill, 1983.
- [39] Shih, Y. and D. Liu, "Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands", *Expert Systems with Applications*, Vol.35, No.1-2, 2008, pp. 350-360.
- [40] Sohn, M., S. Jeong, J. Kim, and H. J. Lee, "Augmented context-based recommendation service framework using knowledge over the linked open data cloud", *Pervasive and Mobile Computing*, Vol.24, 2015, pp. 166-178.
- [41] Uricchio, T., L. Ballan, L. Seidenari, and A. Del Bimbo, "Automatic image annotation via label transfer in the semantic space", *Pattern Recognition*, Vol.71, 2017, pp. 144-157.
- [42] Yazdanfar, N. and A. Thomo, "Link recommender: Collaborative-filtering for recommending URLs to twitter users", *Procedia Computer Science*, Vol.19, 2013, pp. 412-419.
- [43] Zhang, C., Y. Zhang, X. Zhu, Z. Xue, L. Qin, Q. Huang, and Q. Tian, "Socio-mobile landmark recognition using local features with adaptive region selection", *Neurocomputing*, Vol.172, 2016, pp. 100-113.
- [44] Zhao, F., Y. Zhu, H. Jin, and L. T. Yang, "A personalized hashtag recommendation approach using LDA-based topic model in microblog environment", *Future Generation Computer Systems*, Vol.65, 2016, pp. 196-206.

EXIF-based Hashtag Recommender System on Social Networking Service

Sang Hoon Lee* · Su-Yeon Kim**

Abstract

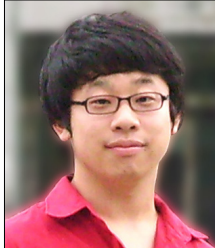
Many users are uploading their daily life activities on SNS and use hashtags to describe their postings. Hashtag has the advantage of letting users specify categories for their postings, however until now, the users has had to manually input the hashtags which has been very inconvenient for them. Therefore, in order to address this issue, this paper proposes a hashtag recommender system which recommends proper hashtags to users based on their uploaded images on SNS. The proposed system is designed using four analytic structures, which is composed of a camera information-based analysis, an address-based analysis, a location based CF analysis, and an image-based analysis. In order to check whether the proposed system is improved compared to the existing systems in terms of the hashtag recommendation function, we conducted an evaluation with 212 SNS users from fifteen countries. As a result of the evaluation process, the proposed system shows very high accuracy recommendation results compared to the existing hashtag recommender systems.

Keywords: Hashtag, Recommender System, EXIF, Social Networking Service

* Researcher, School of Computer & Information Engineering, Daegu University

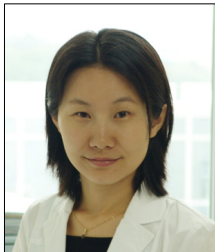
** Corresponding Author, Professor, School of Computer & Information Engineering, Daegu University

◎ 저 자 소 개 ◎



이 상 훈 (prolee@prolee.net)

대구대학교 컴퓨터정보공학부 학사, 석사, 박사 학위를 취득하고 현재 대구대학교 MIS Lab.에서 연구원으로 재직 중이다. 다수 비영리 단체의 서버 구축, 정보시스템 개발 및 관리 경험을 가지고 있다. 주요 관심분야는 추천 시스템, 지식경영, e-비즈니스 등이 있다.



김 수 연 (sykim@daegu.ac.kr)

POSTECH 수학과를 졸업하고, 동 대학교 산업경영공학과 경영정보시스템 전공으로 박사학위를 취득하였다. POSCO ICT와 투이컨설팅에 재직하면서 다수의 시스템 개발 및 SI 컨설팅 프로젝트에 참여하였고, 현재 대구대학교 컴퓨터정보공학부에 재직 중이다. 연구 관심분야로는 추천 시스템, 지식재산권, 기술수용모형, 지식경영 등이 있다.

논문접수일 : 2018년 05월 21일

1차 수정일 : 2018년 08월 14일

게재확정일 : 2018년 10월 08일

2차 수정일 : 2018년 09월 28일