

# 휴먼 FTA를 위한 소셜 웹 마이닝 기반 고용정보 서비스의 설계 및 구현 Design and Implementation of a Employment Information Service based on the Social Web Mining for Human-FTA

송재오, 박용구\*, 유재수  
충북대학교 빅데이터학과,  
어울림컴퍼니(주) 기술연구소\*

Jeo Song, Yong-goo Park\*, Jaesoo Yoo  
Chungbuk National University,  
Research Institute of Eoulimcompany Inc.\*

## 요약

경제혁신 3개년 계획을 토대로 정부는 2015년 국내 생산가능 인구 감소에 대한 대응을 위해 외국인 인력 유치를 위한 휴먼 FTA를 발표하였다. 기존의 외국인 생산 인력에 대한 단순한 양적 증가뿐만이 아니라 해외로 생산거점을 이동한 국내 기업의 리턴을 유도하기 위해 석박사급의 고급 인력과 투자자 유치 등에 대한 내용도 포함하고 있다. 본 논문에서는 상기와 같은 노동시장의 새로운 제도인 휴먼 FTA에 대한 활성화와 원활한 운영을 위해 세계적으로 많이 사용되고 있는 트위터, 페이스북, 구글 등의 소셜 웹 데이터를 활용하여 국내 기업의 외국인 인력에 대한 고용 매칭을 위한 서비스 플랫폼을 제안한다.

## I. 서론

최근 다양한 형태의 취업정보서비스가 인터넷을 통해 이용되고 있지만, 외국인 근로 인력을 위한 일자리 정보 제공에 대하여 특화된 서비스는 미비한 상태이다. 또한, 동남아시아 지역의 일부 국가별 한인단체 또는 여행사 등의 홈페이지를 통해 간단한 구인, 구직 정보가 교류되기는 하지만 국내의 기업들이 통합적으로 외국인 고용 관련 정보를 검색하고 채용할 수 있는 것에는 한계가 있다. 기존의 취업정보서비스는 기업들이 등록하는 구인정보와 기업정보를 구직자들이 검색하여 입사지원을 하는 형태의 단방향적 커뮤니케이션 시스템이 주류를 이루고 있다.

본 논문에서는 트위터, 페이스북, 구글 등의 소셜 웹 데이터를 활용하여 국내 기업의 외국인 인력에 대한 고용 매칭을 위한 HDFS 기반의 서비스 플랫폼을 제안한다. 이를 통해 외국인 구직자가 사용하는 소셜 웹 데이터와 국내 기업들의 구인 정보를 수집하고 분석하여 외국인 구직자에게는 희망 직무에 대한 국내 기업 정보를 SNS를 통해 제공한다. 또한 국내 구인 기업에게는 예비 외국인 근로자의 정보를 제공함으로써 양방향 형태의 커뮤니케이션을 바탕으로 고용서비스가 이루어질 수 있도록 한다.

본 서비스는 국내외에서 많이 사용되고 있는 트위터, 페이스북, 구글 등의 SNS 데이터를 바탕으로 외국인 중심의 구직정보를 수집하고, 취업정보사이트 및 웹 페이지를 통해 국내 기업의 구인정보를 수집한다. 수집된 데이터는 K-평균군집과 텍스트 마이닝을 통해 근로 업무 집중 중심으로 카테고리를 형성하고 분석된다. 여기서 분석된 구인정보는 SNS를 활용하여 외국인 구직자에게 전달이 되고, 구직정보는 웹서비스 플랫폼을 통하여 구인 기업에게 전달이 되는 방식이다. 이와 같은 형태로 매칭이 이루어진 고용의 진행은 서비스 플랫폼을 통해 휴먼 FTA 기준에 부합하는 절차와 서류에 따라 이루어지며, 국내 취업을 위한 온라인 교육콘텐츠 등을 제공한다.



▶▶ 그림 1. 서비스 개요

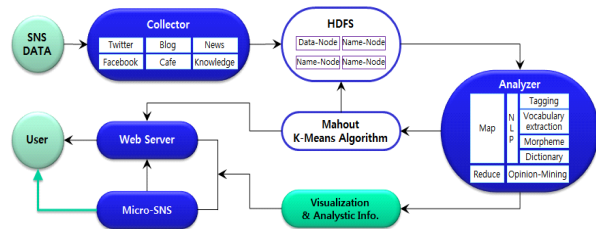
## II. 고용정보 서비스 설계 및 구현

### 1. 서비스 개요

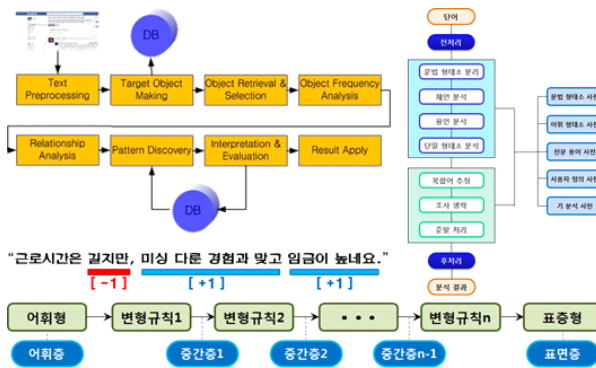
### 2. 시스템 구성

본 논문에서 제안하는 시스템은 아래 그림 2. 와 같이

크게 소셜 웹 데이터 수집을 위한 Collector, 고용정보 분석과 서비스 분류를 위한 Analyzer, 정보서비스 제공을 위한 Visualization&Analytics Info. 의 세가지 부분으로 구성되어 있다. 서비스 이용자는 웹을 통한 플랫폼 서비스를 기존의 취업정보사이트와 같이 이용할 수 있으며, 휴먼 FTA를 활용한 고용 진행은 해당 플랫폼 시스템만을 통해 가능하다. Micro-SNS 는 트위터, 페이스북 등을 통하여 외국인 구직자가 국내 기업의 구인 정보를 제공받을 수 있도록 하는 부분으로 Micro-Targeting 형태의 서비스 성격을 가지고 있다. 따라서 서비스 카테고리는 국내 기업의 구인 정보를 기준으로 분류되기 위하여 분석되며, 외국인의 구직정보는 국가별로 성별, 관심업무, 보유 능력 및 경력과 학력 등을 위해 분석된다.



▶▶ 그림 2. 시스템 구성



▶▶ 그림 3. 자연어 처리

그림 3은 자연어 처리에 관한 내용으로 촘스키의 변형 생성이론 모델을 기초로 구현하였으며 정확성을 위해 어휘사전, 전문용어사전, 은어사전을 함께 사용하였다. 글로벌 데이터 처리를 위한 영문 데이터에 대한 분석은 스탠포드에서 제공하는 Part-of-Speech(POS) Tagger와 SentiWordNet의 감성사전을 이용해 구현하였다.

근로 직종에 대한 카테고리 분류를 위한 군집 분석은 Mahout의 K-평균군집을 위한 EuclideanDistanceSimilarity 클래스를 이용하여 외국인 구직자간의 유사도를 측정하여 나타내었다.

$$V = \sum_{i=1}^k \sum_{j \in S_i} [x_j - \mu_i]^2$$

표 1. EuclideanDistanceSimilarity 사용 유사도 측정

	서비스	건설	생산	거리	유사도
외국인력1	3.0	2.5	5.0	0.000	1.000
외국인력2	2.5	5.0	2.0	3.937	0.203
외국인력3	-	-	2.5	2.500	0.286
외국인력4	-	3.0	5.0	0.500	0.667
외국인력5	3.0	2.0	4.0	1.118	0.472

### III. 결론

본 논문에서는 트위터, 페이스북, 구글 등의 소셜 웹 데이터 마이닝과 HDFS 등의 분산처리시스템을 활용하여 방대한 양의 글로벌 구인, 구직 정보를 분석하여 휴먼 FTA에 활용하는 서비스에 대하여 연구하고자 하였다. 국내 리턴 기업을 증가시키고, 외국인 인력에 대한 제도적인 시스템을 위해 영문 데이터 처리도 사용하였으나, 실제로 국내에서 근로하는 외국인의 출신 국가를 고려한다면 지속적인 언어 확장이 필요하며, 국내 사용자를 위한 한글 형태소 분석기에 대한 개선이 필수적이다. 실제 본 연구에서 구현한 시스템은 한글 형태소 분석기의 정확도를 위해 사전 기능을 계속적으로 사용하고 추가하였다. 이와 같은 시스템적 개선과 고용노동부에서 사용하는 기존의 외국인 고용관리시스템(EPS) 연계가 가능하다면, 휴먼 FTA 활성화를 위한 보다 실질적이고 효율적인 운영 및 관리 시스템이 될 수 있다고 기대한다.

### ■ 참고 문헌 ■

- [1] “노동시장 이중구조를 완화하고 휴먼 FTA 등을 통한 생산인력 확충”, 2015년 경제정책방향, 경제혁신 3개년 계획 관계부처 합동 보고서, pp.15-16, 2014.
- [2] 송재오, 이상문, “K-평균 군집을 이용한 마이크로타겟팅을 위한 SNS 빅데이터 활용 모델링에 관한 연구”, 한국컴퓨터정보학회 동계학술대회 논문집, 제22권, 제1호, pp.321-324, 2015.
- [3] Jeo Song, Da Young Yun, Jae Moon Kim, Sang Moon Lee, “Cloud-based Enterprise Information Portal System Using SNS Big Data for Business Competitiveness”, MITA international Conference, HongKong, Vol.10, 2014.
- [4] Sean Owen, Robin Anil, Ted Dunning, Ellen Friedman, MAHOUT IN ACTION, Manning Publications, USA, 2011.
- [5] Jeffrey Dean, Sanjay Ghemawat, “MapReduce: a flexible data processing tool”, InProc of Magazine Communications of the ACM, Vol.53, Issue 1, January, 2010.