Journal of the Korean Society of Marine Engineering, Vol. 39, No. 1 pp. 58~62, 2015
J. Korean Soc. of Marime Engineering (JKOSME)
http://dx.doi.org/10.5916/jkosme.2015.39.1.58

ISSN 2234-7925 (Print)
ISSN 2234-8352 (Online)
**Original Paper**

# 다양한 어휘 가중치를 이용한 블로그 포스트의 자동 분류

김수아[1] · 조희선[2] · 이현아†

(Received July 28, 2014 ; Revised September 22, 2014 ; Accepted December 19, 2014)

# Automatic Classification of Blog Posts using Various Term Weighting

Su-Ah Kim[1] · Hee-Sun Jho[2] · Hyun Ah Lee†

**요약:** 대부분의 블로그 사이트에서는 미리 정의된 분류 체계에 따른 내용 기반 분류 환경을 제공하고 있으나, 작성된 포스트의 분류를 수동으로 선택해야하는 번거로움 때문에 대부분의 블로거들은 포스트에 대한 분류를 입력하지 않고 있다. 본 논문에서는 블로그 포스트의 자동 분류를 위해 블로그 사이트에서 분류별 문서를 수집하고 수집된 분류별 문서의 어휘빈도와 문서빈도, 분류별 빈도 등의 다양한 어휘 가중치 조합하여 블로그 포스트의 특성에 적합한 가중치 방식을 찾고자 한다. 실험에서는 본 논문에서 제안한 TF-CTF-IECDF를 어휘 가중치로 사용한 분류 모델이 77.02%의 분류 정확률을 보였다.

**주제어:** 블로그 포스트, 자동 분류, 어휘 가중치, 가중치 조합

**Abstract:** Most blog sites provide predefined classes based on contents or topics, but few bloggers choose classes for their posts because of its cumbersome manual process. This paper proposes an automatic blog post classification method that variously combines term frequency, document frequency and class frequency from each classes to find appropriate weighting scheme. In experiment, combination of term frequency, category term frequency and inversed (excepted category's) document frequency shows 77.02% classification precisions.

**Keywords:** Blog post, Automatic classification, Term weighting, Weighting combination

## 1. 서 론

정보 개방과 사용자 참여를 중심으로 한 웹 2.0의 시대를 맞이하여 블로그는 1인 미디어로 급부상하고 있다. 개인적 기록이나 사회 참여의 도구로 주로 사용되던 블로그가 최근에는 취미나 관심 분야의 정보 획득 및 공유의 목적으로 많이 사용하고 있다**[1]**. 정보 획득을 위한 검색에서 각광받던 지식iN 등의 지식 서비스가 스팸을 포함한 부정확한 답변의 문제가 커지면서, 블로그와 카페글에서 제공되는 다양한 전문 정보들의 활용이 높아지고 있는 실정에서, 포털 사이트의 검색 결과에서 블로그 정보를 최상단에 노출하는 것은 블로그의 정보 제공 기능을 입증하는 것으로 볼 수 있다.

블로그 사용자가 늘어남에 따라서 블로그 서비스를 제공하는 사이트에서는 주제와 목적에 맞게 블로그나 블로그 포스트를 분류하여, 블로그를 통한 정보 획득에 필요한 시간과 노력을 줄일 수 있게 지원한다. 하지만, 한 블로거가 다양한 분야(예를 들어 영화, IT, 주식)의 포스트를 작성하는 블로그의 특성상, 한 블로그를 하나의 분류로 대응시키기 적합하지 않을 수 있다. 네이버 블로그의 경우 포스트를 작성할 때마다 내용에 적합한 분류를 사용자가 입력할 수 있게 하여 포스트별 분류를 제공하지만, 대부분의 블로거들이 기본 분류를 선택하는 양상을 보이고 있다. **Figure 1**은 블로그 포스트의 예를 보인다. 해양 과학에 관련된 내용을 다루고 있으나, '좋은 완구' 카테고리에 포함되어 있으며,

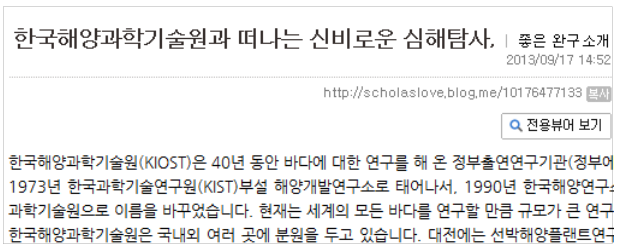


**Figure 1:** An example of a mis-classified blog post

† Corresponding Author (ORCID: http://orcid.org/0000-0001-6753-825X): Department of Computer Software Engineering, Kumoh National Institute of Technology, 61 Daehak-ro, Gumi, Gyeongbuk, 730-701, Korea, E-mail: halee@kumoh.ac.kr Tel: 054-478-7546
1 Department of Computer Software Engineering, Kumoh National Institute of Technology, E-mail: sa4956@nate.com, Tel: 054-478-7546
2 Department of Computer Software Engineering, Kumoh National Institute of Technology, E-mail: shinhwa3528@naver.com, Tel: 054-478-7546

적절한 분류도 선택되어 있지 않아, 해당 내용에 관심있는 사용자들이 참조하기 어려운 문제를 가지고 있다. 이러한 경우 작성한 블로그의 내용에 맞는 분류를 자동으로 추천하는 기능이 제공된다면 블로그 포스트의 활용성이 증대될 수 있다.

문헌에 나타난 용어의 특성을 분석하면 문헌의 자동 분류가 가능하며, 이는 블로그 포스트에도 적용가능하다. 국내 연구 중 **[2]**에서는 주제 분별 용어 집합을 생성하고 주제 분별 점수를 계산한 뒤 다양한 분류 모델을 사용하였다. 분류 체계가 명확한 신문 기사에 대한 실험 결과에서는 높은 성능을 보였으나, 클래스간의 연관성이 높은 경우에는 낮은 성능을 보이는 문제점이 있어 추가적인 학습 문서의 정제가 필요하였다. **[3]**에서는 시간대별로 블로그의 주제 특성이 있다는 점에 착안하여, 단어 빈도를 가중치로 하여 얻은 문서 간 코사인 유사도를 이용하여 계층적 클러스터를 구성하되, 시간 주기성을 반영하여 대표 클러스터를 결정하는 방법을 제안하고 있으나, 이 방식은 내용이나 분류별 특징을 충분히 반영하지 못한다. 국외 연구의 경우 블로그 자체를 분류하는 다양한 시도**[4]-[6]**는 존재하였으나 블로그 포스트를 내용 기반으로 분류하는 연구는 많지 않다. **[7]**에서는 영역별 사전을 이용하여 포스트에 대한 내용 분류를 시도하였으나 수동 구축한 사전에 기반하고 있으며, **[8]**에서는 위키피디아에서 추출한 정보를 이용한 블로그 포스트 분류를 시도하고 있으나 한정적인 영역에 대한 높지 않은 성능을 보이고 있다. **[9]**에서는 TF-IDF를 가중치 기법으로 사용하여 나이브 베이지안과 ANN분류 방식을 이용하여 블로그 포스트에 대한 지도 학습에 기반한 분류 방식을 보인다.

본 논문에서는 네이버 블로그에서 주제별 분류가 등록된 포스트들을 수집하고, 이를 학습 데이터로 사용하여 자동으로 포스트의 주제별 분류를 추천하기 위한 시스템을 제안한다. 시스템에서는 TF-IDF 이외의 다양한 가중치 기법을 제안하고 평가하여, 블로그 분류에 적합한 방식을 찾고자 한다. 또한 다양한 분류기를 적용하여 포스트에 맞는 분류를 결정한다.

## 2. 블로그 포스트 자동 분류 시스템

본 논문에서 제안하는 시스템은 블로그 포스트를 수집한 뒤, 분류를 위해 제안된 다양한 어휘 가중치를 여러 분류 모델에 적용하여 포스트의 분류를 결정한다. 아래에서는 각 단계에 대해 설명한다.

### 2.1 블로그 포스트 수집

국내의 대표적인 블로그 사이트인 네이버, 다음, 티스토리에서 수집된 블로그 문서를 기준으로 적합한 분류를 조사한 연구**[10]**에서 네이버 블로그는 장르 분류 일치도에서도 높은 결과를 보였다. 본 논문에서는 주제 분류가 부착된 네이버 블로그 포스트를 이용하여 학습에 사용한다.

네이버 블로그에는 **Figure 2**의 왼쪽에서 보이는 30개의 분류가 존재한다. 이 중 일부 분류들은 학습 데이터로 쓸 만큼 충분한 양의 글이 올라오지 않거나, 분류의 주제에 맞지 않는 광고성 글이 높은 비율을 보인다. 이로 인한 문제를 보완하기 위해 네이버의 30개의 분류 중 일부를 제거하거나 병합하여 **Figure 2**의 오른쪽과 같은 16개 분류를 얻고, 이를 이용하여 자동 분류를 수행한다.

### 2.2 단어별 주제 분별 점수 계산

문서 분류를 위한 문서 특성은 제목과 본문에서 사용되는 명사에서 추출한다. 한국어 형태소 분석기를 이용하여 문서 내의 단어를 추출한 뒤, 각 단어의 빈도를 분석한다. 본 논문에서는 기존의 방식에서 주제 분별력을 파악하기 위해서 이용하는 TF와 IDF에 추가하여, 군집 내에서의 단어 빈도나 문서 빈도를 반영하는 CTF, CDF, IECDF를 단어 가중치로 제안한다. 아래에서는 각각에 대해 설명한다.

#### 2.2.1 TF와 IDF

TF (Term Frequency)는 각 문서에서의 단어 빈도로 단어 가중치를 계산한다. 문서의 크기가 커지면 문서에서 발생한 단어의 절대 빈도도 커지므로, 문서 D에서 발생한 단어 $w_i$의 빈도수 $freq(w_i, D)$에 문서 D의 총 단어수로 나누어, 정규화한 $TF_D(w_i)$를 **Equation (1)**로 구한다.

$$TF_D(w_i) = \frac{freq(w_i, D)}{\sum_{w_j \in D} freq(w_j, D)} \quad (1)$$

IDF (Inversed Document Frequency)는 문서 빈도 역수를 이용하여 단어의 희소성이나 정보성을 표현하는 통계적 방법이다. 분류 C에 속하면서 단어 $w_i$를 포함하는 문서를

| Entertainment/Art | Living/Knowhow/Shopping | Hobby/Leisure/Travel | Knowledge/Trend |
|---|---|---|---|
| Culture·Book<br>Movie<br>Art·Design<br>Concert·Exhibition<br>Music<br>Drama<br>Star·Entertainer<br>Cartoon·Animation<br>Broadcast | Ordinary·Thinking<br>Child Card·Wedding<br>Pet·Companion Animal<br>Book Note·Image<br>Fashion·Beauty<br>Interior·DIY<br>Cooking·Recipe<br>Product Review | Game<br>Sports<br>Photo<br>Car<br>Hobby<br>Domestic Travel<br>Oversea Travel<br>Restaurant | IT·Computer<br>Social·Politics<br>Health·Medical<br>Business·Economy<br>Language<br>Education·Studies |

➡

| Culture.Book | Art.Design/Concert.Exhibition |
|---|---|
| Movie | Cook.Recipe.Restaurant |
| Music | Domestic/Oversea travel |
| Sports | Pet.Companion Animal |
| Game | IT·Computer |
| Interior | Child Card·Wedding |
| Car | Fashion·Beauty |
| Social.Politics | Health·Medical |

**Figure 2:** Original classes of Naver blog post and our simplified blog classes

$D_{w_i,C}$로 표기하고, 이 문서의 개수를 $|D_{w_i,C}|$로 표기하자. $IDF(w_i)$는 분류와는 상관없이, 전체 문서수 $|D_{all,all}|$를 단어 $w_i$가 발생한 문서의 빈도 $|D_{w_i,all}|$를 나눈 값에 log를 취하여 **Equation (2)**와 같이 구한다.

$$IDF(w_i) = \log \frac{|D_{all,all}|}{|D_{w_i,all}|} \tag{2}$$

2.2.2 CTF

주제 분별력이 높은 단어는 특정 분류 즉 대표 분류에서만 자주 발생할 것으로 기대할 수 있다. 이는 분류 내 단어 빈도 CTF (Category Term Frequency)로 수치화할 수 있다. 본 논문에서는 **Equation (3)**과 **Equation (4)**로 CTF를 구한다. **Equation (3)**에서는 단어의 누적 빈도가 가장 높은 분류를 대표 분류로 보고, 단어 $w_i$의 대표 분류 $Max\,C_{w_i}$를 구한다. 얻어진 대표 분류에서의 단어 $w_i$의 누적빈도 $CTF(w_i)$를 **Equation (4)**로 얻는다. 시스템에서는 학습과정에서 각 분류별 문서개수를 동일하게 구성하여, CTF에 대한 분류별 정규화가 이루어지도록 한다.

$$Max\,C_{w_i} = \arg_C \max \sum_{D \in C} freq(w_i, D) \tag{3}$$

$$CTF(w_i) = \log \sum_{D \in Max\,C_{w_i}} freq(w_i, D) \tag{4}$$

2.2.3 CDF

주제 분별력이 높은 단어는 해당 주제의 대부분의 문서에 나타날 것으로 기대할 수 있다. 이는 대표 분류에서의 문서 빈도 CDF (Category Document Frequency)로 수치화할 수 있다. **Equation (5)**은 **Equation (3)**에서 구한 단어 $w_i$의 대표 분류 $Max\,C_{w_i}$에서의 문서빈도인 $CDF(w_i)$를 구한다. 식에서는 단어 $w_i$가 발생한 대표 분류 내 문서 개수 $|D_{w_i,Max\,C_{w_i}}|$를 대표 분류의 전체 문서 개수인 $|D_{all,Max\,C_{w_i}}|$로 나누어, 단어 $w_i$가 분류의 대표성을 얼마나 띄고 있는냐를 수치화한다.

$$CDF(w_i) = \frac{|D_{w_i,Max\,C_{w_i}}|}{|D_{all,Max\,C_{w_i}}|} \tag{5}$$

2.2.4 IECDF

대표 분류를 제외한 분류의 문서들이 그 단어를 덜 포함할수록 대표 분류에 대한 주제 분별력이 높다고 볼 수 있다. 이는 대표 분류 이외 문서에서의 IDF인 IECDF (Inversed, Excepted Category's, Document Frequency)로 수치화할 수 있다. **Equation (6)**는 IECDF를 구한다. 수식에서 $\overline{Max\,C_{w_i}}$는 전체 분류에서 **Equation (3)**의 $Max\,C_{w_i}$를 제외한 모든 분류를 의미한다. 대표 분류 이외의 분류에서의 문서 빈도를 분모로 사용하여, 대표 분류 이외의 문서에서 적게 나타날수록 높은 단어의 가중치를 얻는다.

$$IECDF(w_i) = \log \frac{|D_{all,\overline{Max\,C_{w_i}}}|}{|D_{w_i,\overline{Max\,C_{w_i}}}|} \tag{6}$$

### 2.3 주제 분별력 점수 결합

본 논문에서는 위에서 제안한 가중치를 다양하게 조합하여 단어의 주제 분별 점수를 구한다. 조합에서는 일곱 가지 방식을 사용하며, 각 조합에서는 2.2에서 제안한 단어가중치의 곱으로 가중치를 결합한다.

첫 번째 방식으로 기존의 일반적인 가중치 기법인 TF-IDF를 사용한다. CTF와 CDF의 유용성을 확인하기 위해 두 번째 결합으로 TF-CTF, 세 번째 결합으로 TF-CDF를 사용한다. 네 번째 결합으로는 역문서 빈도의 성격인 IDF와 IECDF를 쓰지 않는 TF-CTF-CDF를 사용한다. 다섯 번째 결합으로 TF-CTF-IECDF를 사용한다. TF와 CTF, IECDF를 곱하여 대표 분류에서의 빈도와 이외 분류에서의 IDF를 반영하여 주제 분별 점수를 계산한다. 여섯 번째 결합으로는 TF-CDF-IDF를 사용하여, CDF를 통해 해당 단어가 대표 분류에서 폭넓게 사용될수록, IDF를 통해 해당 단어가 희소성이 높을수록 높은 점수를 얻도록 한다. 마지막 결합 방식으로는 TF-CDF-IECDF는 IDF 대신 IECDF를 사용하여 나머지 분류에서의 희소성이 높을수록 높은 점수를 얻도록 한다.

### 2.4 분류 모델 생성

문서별로 주제 분별 점수가 구해지면 이를 이용하여 분류 모델을 생성한다. 본 논문에서는 분류 모델 생성을 위하여 기존의 소프트웨어 WEKA 3.6.10**[11]**에 구현된 Complement Naive Bayes와 Naive Bayes Multinomial 알고리즘을 사용하였다.

각 분류 알고리즘을 이용해 각 문서별 용어에 대한 주제 분별 점수를 입력으로 하고, 결과 분류를 출력으로 설정하여 각각의 분류모델을 생성하였다. 분류 모델의 검증은 생성과 마찬가지로 검증용 문서 집합을 이용하여 생성된 각 분류 모델의 정확도를 검증한다.

## 3. 실험 및 평가

본 논문에서 제안하는 가중치 결합 방식과 2가지의 분류 학습기를 통한 분류 정확률을 평가하였다. 실험에서는 16개의 각 분류에서 임의로 추출한 500개, 총 8000개의 학습 데이터를 이용하였으며, 실험 데이터는 분류별 200개,

총 3200개를 사용하였다.

**Table 1**은 결과를 보인다. 기존의 방식에서 이용하던 TF와 IDF를 결합한 방식은 50%가 되지 않는 분류 정확률을 보였다. TF-IDF는 분류 정보가 반영되지 않고, 키워드의 단순 빈도와 전체 문서 집합에서의 IDF를 이용하여 문서를 분류하여 낮은 정확률을 보이는 것으로 분석되었다. 이에 비하여 분류 정보가 사용된 TF-CTF와 TF-CDF는 이보다 높은 70% 내외의 성능을 보여 분류에 기반한 정보의 효과를 확인할 수 있었다.

**Table 1**: Precision for each classifier

| | Complement Naive Bayes | Naive Bayes Multinomial |
|---|---|---|
| TF-IDF | 44.06% | 43.31% |
| TF-CTF | 75.59% | 71.95% |
| TF-CDF | 67.70% | 63.60% |
| TF-CTF-CDF | 66.64% | 63.76% |
| TF-CTF-IECDF | 75.03% | 77.02% |
| TF-CDF-IDF | 70.01% | 70.29% |
| TF-CDF-IECDF | 72.40% | 72.61% |

결과에서 Naive Bayes Multinomial로 학습을 한 결과에서 TF-CTF-IECDF가 77.02%로 가장 높은 정확률을 보였다. 전체 실험 결과에서 역문서빈도인 IDF를 사용한 방식보다 대표 분류 외 분류에서의 역문서빈도인 IECDF가 좋은 성능을 보였다. 또한 CDF보다 CTF가 좋은 성능을 보여, 블로그 포스트 분류에서는 포함된 문서 개수보다는 단어 발생의 중복성이 고려되는 단어빈도가 유용함을 알 수 있었다.

오류 분석에서는 포스트의 분류 모호성, 정보성이 떨어지는 외래어에 의한 오류, 이슈가 되는 고유 명사에 의한 오류 등이 주요한 문제로 나타났다. 포스트의 분류 모호성에서는 Game과 Sports 두 분류에 연관된 e-sports 대회 관련 포스트, Book과 Child Care에 연관된 육아 서적 관련 포스트, Book과 Movie에 연관된 원작 소설에 기반한 영화 관련 포스트, Car와 Game에 연관된 자동차 레이싱 게임 관련 포스트 등과 같이, 두 분류에 모두 속할 수 있는 포스트에 의한 오류가 나타났다. 외래어에 의한 오류는, 영어 가사에서 발생하는 불용어들(예를 들어, the, to, and, in, is 등)이 네이버 블로그 포스트에서는 IDF나 CDF, IECDF가 높게 계산되는 등의 문제로 나타났으며, 외래어 불용어에 대한 별도의 처리가 필요한 것으로 분석되었다. 이슈가 되는 고유 명사에 의한 오류에서는, 예를 들어 “은밀하게 위대하게”가 개봉한 직후에 “김수현”, “은밀”, “웹툰” 등이 단기간에 Movie 분류에 집중적으로 발생하여 분류 정확도에 악영향을 미쳐, 이슈 단어나 단어 발생 분포에 대한 별도의 고려도 필요한 것으로 나타났다.

본 논문에서는 실험 결과에서 정확률이 높은 상위 6개를 이용하여 블로그 자동 분류 시스템을 구축하였다. **Figure 3**은 실행 예를 보인다. 시스템에서 사용자가 포스트를 작성한 뒤 저장 버튼을 누르면 자동으로 추천 카테고리를 제시하는 방법으로 구동된다. 시스템에서는 정확률이 높은 6개의 분류 기법에서 1위 분류에 투표 방식(voting)을 적용하여, 가장 많이 추천된 분류부터 순서대로 사용자에게 제시한다. 실행 예에서 포스트는 육아와 관련한 책을 소개하는 글로서, [문학-책]과 [육아_결혼]의 분류를 추천하는 결과를 볼 수 있다.

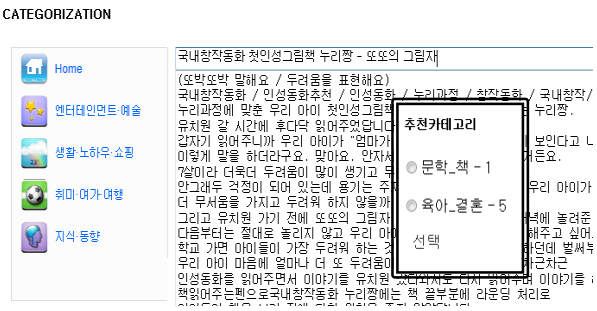


**Figure 3:** An Example of class recommendation for an input post

## 4. 결 론

문서 내에서 중요한 키워드를 뽑아내는 TF-IDF는 다양하게 변형되어 문서 분류에도 활용되고 있다. 블로그 포스트의 정보성이 높아지고 있지만, 기존의 문서 분류의 연구는 뉴스 기사와 같이 의도가 분명한 글에 관한 분류에 집중되어 있어, 다양한 개성을 가진 사용자들이 작성하는 블로그 포스트에 적용하기 적합하지 않다.

본 논문에서는 TF-IDF를 변형하여 블로그 포스트를 자동으로 분류하기 위해 단어 주제 분별력을 계산하기 위한 다양한 가중치를 제안하였다. TF와 IDF를 각각 카테고리로 확장시킨 개념의 CDF와 CTF, IECDF에 대한 실험에서는, IDF보다 IECDF가 블로그 문서의 분류의 정확도를 높였으며, 단순 단어 빈도 TF보다 분류로 확장한 단어 빈도나 문서빈도인 CTF나 CDF가 더 정확한 결과를 보였다.

블로그 문서를 자동으로 분류하는 데는 한계가 있을 수 있다. 정형적인 텍스트가 아니기 때문에 오타나 신조어 등에 민감할 수 있는데 이러한 점은 형태소 분석기의 성능이 향상되거나 고유 명사 사전 등을 구축하면 해결 할 수 있을 것이라고 기대한다. 추후 연구로는 문서 자동 분류를 문서 필터링으로 확장시킬 예정이다.

## 후 기

이 연구는 금오공과대학교 학술연구비에 의하여 지원된 논문

## References


[1] Y. J. Kim, “A study on the blog as a media : Focused on media functions and the problems of the blog,” Korean Journal of Journalism & Communication Studies, vol. 50, no. 2, pp. 59-90, 2006 (in Korean).

[2] D. H. Park, W. S. Choi, and H. J. Kim, “Web document classification based on hangeul morpheme and keyword analyses,” Transactions of the Korean Information Processing Society Transaction : Part D (Database), vol. 19-D, no. 4, pp. 263-270, 2012 (in Korean).

[3] S. W. Lee, D. J. Choi, H. W. Jung, and J. H. Lee, “Study of blog auto categorizing based on time periodicity,” Proceedings of Korean Institute of Intelligent Systems Spring Conference, vol. 21, no. 1, pp. 86-87, 2011 (in Korean).

[4] H. Qu, A. L. Pietra, and S. Poon “Automated blog classification: challenges and pitfalls,” Association for the Advancement of Artificial Intelligence Spring Symposium : Computational Approaches to Analyzing Weblogs, pp. 184-186, 2006.

[5] D. Ikeda, H. Takamura, and M. Okumura, “Semi-supervised learning for blog classification,” Proceedings of the 23th Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, vol. 2, pp. 1156-1161, 2008.

[6] E. Lex, C. Seifert, M. Cranitzer, and A. Juffinger, “Automated blog classification : A cross domain approach,” Proceedings of the International Association for Development of the Information Society, International Conference on WWW/Internet, p. 598, 2009.

[7] C. Hashimoto and S. Kurohashi, “Blog categorization exploiting domain dictionary and dynamically estimated domains of unknown words,” Proceedings of ACL-08, HLT Short Papers, pp 69-72, 2008.

[8] Stephanie D. Husby and Denilson Barbosa, “Topic classification of blog posts using distant supervision,” Proceedings of the 13th Conference of the European Chapter of Association for Computational Linguistics, pp 28-36, 2012.

[9] M. K. Dalal and M. A. Zaveri, “Automatic classification of unstructured blog text,” Journal of Intelligent Learning Systems and Applications, vol. 5, no. 4, pp. 108-114, 2013.

[10] H. Y. Kim, An Experimental Study on Semi-Supervised Classification of Blog Genres, MS Thesis, Yonsei University, Korea, 2009 (in Korean).

[11] http://www.cs.waikato.ac.nz/ml/weka/, Accessed July 25, 2014.

[12] S. A. Kim, H. S. Cho, and H. A. Lee, “Automatic classification of blog posts,” Technology of the 25th Annual Conference on Human and Cognitive Language, pp. 160-162, 2013 (in Korean).